From THE DEPARTMENT OF CLINICAL NEUROSCIENCE

Karolinska Institutet, Stockholm, Sweden

STUDIES OF RISK LOCI IN MULTIPLE SCLEROSIS

Magdalena Lindén

Karolinska Institutet

Stockholm 2014

Published by Karolinska Institutet. Printed by åtta.45 Tryckeri AB, Solna, Sweden.

ISBN 978-91-7549-558-3

Studies of Risk Loci in Multiple Sclerosis

THESIS FOR DOCTORAL DEGREE (Ph.D.)

by

Magdalena Lindén

Principal Supervisor:
Associate Professor Ingrid Kockum
Karolinska Institutet
Department of Clinical Neuroscience

Co-supervisors:
Professor Tomas Olsson
Karolinska Institutet
Department of Clinical Neuroscience

Dr Mohsen Khademi
Karolinska Institutet
Department of Clinical Neuroscience

Opponent:
Associate Professor David Booth
University of Sydney
Westmead Millenium Institute

Examination Board:
Associate Professor Erik Melén
Karolinska Institutet
Institute of Environmental Medicine

Professor Annika Lindblom
Karolinska Institutet
Department of Molecular Medicine and Surgery

Professor Oluf Andersen
Göteborgs Universistet
Institute of Neuroscience and Physiology

“Our minds possess by nature an insatiable desire to know the truth.”

Cicero (106-43 B.C.), Tusculanarum Disputationum I, 18

ABSTRACT

Multiple sclerosis (MS), a chronic inflammatory disease of the central nervous system, is one of the most common causes of neurological disability among young adults. While the etiology is unknown, an increasing number of environmental and genetic risk factors is being identified. The major contribution to the genetic risk in MS is by alleles of *HLA* genes, with the *HLA-DRB1*15:01* allele being the strongest risk factor. At present also more than 100 non-*HLA* loci have been identified.

The main objective of the research presented in this thesis was to study genetic variants and their contribution to MS risk.

In **Paper I** we investigated whether single nucleotide polymorhisms (SNPs) in the region including the *IL21* gene, associated with other inflammatory diseases, also contribute to an increased risk for MS. Our results confirmed that this locus does most probably not have a strong effect, if any, on MS risk.

In **Paper II** we studied protein levels of the CXCL13 chemokine as measured in the cerebrospinal fluid in relation to a number of MS-associated genetic variants. We found that the risk genotypes of *HLA-DRB1*15:01* as well as SNPs in the regions of the *IRF5, OLIG3/TNFAIP3* and *RGS1* genes were associated with increased levels of CXCL13 which has also been suggested as a biomarker for a more severe disease course of relapsing-remitting MS.

In **Paper III** we performed an analysis of gene-gene interactions between the currently established risk loci in MS and found three pairs of interacting variants with significant interaction effects that departed from additivity of the separate effects of the single variants. We found interactions between *HLA-A*02:01* and *HLA-DRB1*15:01,* the risk allele of rs6677309 and *HLA-DRB1*15:01,* as well as the risk allele of rs7196953 and *HLA-DRB1*03:01*. The findings suggest that these pairs of risk factors are involved in the same sufficient causes for MS.

In **Paper IV** we studied *cis*-regulation of gene expression by MS associated loci and found several expression quantitative loci in peripheral blood mononuclear cells. We replicated some of them in lymphoblastic cell lines, and observed that one of the eQTLs was also active in $CD19^+$, and another in $CD4^+$ cells.

Replications and functional experiments are needed in order to take the results presented in this thesis further in the search for biological pathways involved in MS disease mechanisms.

LIST OF SCIENTIFIC PAPERS

I. **No evidence of *IL21* association with multiple sclerosis in a Swedish population**
LINDÉN M, Nohra R, Sundqvist E, Khademi M, Hillert J, Alfredsson L, Olsson T, Kockum I
Tissue Antigens 2011 Oct;78(4):271-4

II. **Multiple sclerosis risk genotypes correlate with an elevated level of the suggested prognostic biomarker CXCL13**
LINDÉN M, Khademi M, Lima Bomfim I, Piehl F, Jagodic M, Kockum I[*], Olsson T[*]
Multiple Sclerosis Journal 2013 Jun;19(7):863-70
[*]Authors contributed equally

III. **Genetic interaction analysis of multiple sclerosis risk loci**
LINDÉN M, Lima Bomfim I, Hillert J, Alfredsson L, Olsson T, Kockum I
Manuscript

IV. **Impact of genetic risk loci in multiple sclerosis on expression of proximal genes**
James T[*], LINDÉN M[*], Huss M, Brandi M, Khademi M, Tegnér J, Gomez-Cabrero D, Kockum I[‡], Olsson T[‡]
Manuscript
[*],[‡] Authors contributed equally

CONTENTS

1 Introduction 1
1.1 Multiple sclerosis 1
1.1.1 Epidemiology 1
1.1.2. Clinical manifestations 1
1.1.3 Immunopathology 3
1.1.4 Treatments 4
1.2 Etiological studies of multiple sclerosis 4
1.2.1 Environmental factors 5
1.2.2 Genetics 5
1.2.2.1 Genetic variation and single nucleotide polymorphisms (SNPs) 6
1.2.2.2 Linkage disequilibrium 6
1.2.2.3 The HLA region 7
1.2.2.4 Genome wide association studies (GWAS) 8
1.2.2.5 The ImmunoChip study 9
1.2.2.6 Missing heritability and gene-gene interactions 10
1.2.3 From genetic variation to function 11
1.2.3.1 Transcriptomics and eQTL analyses 11
1.2.3.2 Biomarkers 12
1.2.3.3 Bioinformatics and integration of functional data 13
2 Aims of the thesis 14
3 Methodological considerations 15
3.1 Cases and controls 15
3.1.1 EIMS 15
3.1.2 IMSE 15
3.1.3 BRAIN 15
3.1.4 GEMS 15
3.1.5 STOP-MS 16
3.2 Genotyping 16
3.2.1 HLA genotyping 16
3.2.2 HLA imputation 16
3.2.3 SNP genotyping 17
3.3 Enzyme-linked immunosorbent assay (ELISA) 17
3.4 Gene expression analysis 18
3.4.1 Real-time quantitative RT-PCR 18
3.4.2 RNA sequencing 18

3.5 Statistical analyses 19
3.5.1 Case-control association tests 19
3.5.2 Meta-analysis 19
3.5.3 Quantitative trait association analysis 19
3.5.4 Interaction analysis 19
3.5.5 Generalized linear model for eQTL analysis 19
3.5.6 Multiple comparisons, permutations and bootstrapping 20
4 Results and Discussion, ... 21
4.1 Paper I 21
4.2 Paper II 22
4.3 Paper III 26
4.4 Paper IV 29
4.5 Summary of findings 33
5 Concluding remarks 34
6 Future perspectives 36
7 Acknowledgements 38
8 References 41

LIST OF ABBREVIATIONS

APC	Antigen-presenting cell
BBB	Blood-brain barrier
CD	Cluster of differentiation
cDNA	Copy DNA
CI	Confidence interval
CIS	Clinically isolated syndrome
CNS	Central nervous system
CSF	Cerebrospinal fluid
CXCL13	C-X-C motif chemokine 13
DNA	Deoxyribonucleic acid
EAE	Experimental autoimmune encephalomyelitis
EBV	Epstein-Barr virus
ELISA	Enzyme-linked immunosorbent assay
eQTL	Expression quantitative trait locus
FDR	False discovery rate
GWAS	Genome-wide association study
HLA	Human leukocyte antigen
iOND	Inflammatory other neurological diseases
kb	Kilobases
LCL	Lymphoblastic cell line
LD	Linkage disequlibrium
MALDI-TOF	Matrix-assisted laser desorption/ionization time-of-flight
MRI	Magnetic resonance imaging
mRNA	Messenger RNA
MS	Multiple sclerosis
OCB	Oligoclonal bands
OND	Other neurological diseases
OR	Odds ratio
PBMC	Peripheral blood mononuclear cell
PCA	Principal component analysis
PCR	Polymerase chain reaction
PPMS	Primary progressive MS
qRT-PCR	Quantitative reverse transcription PCR
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
RRMS	Relapsing-remitting MS
SD	Standard deviation
SNP	Single nucleotide polymorphism
SPMS	Secondary progressive MS

1 INTRODUCTION

1.1 MULTIPLE SCLEROSIS

The pathological changes found in multiple sclerosis (MS) were first described in 1868 by the French neurologist Jean-Martin Charcot, who identified lesions in post-mortem brains and spinal cords of patients with this condition, and therefore called it “la sclérose en plaques” (1). Today, this disease is known to be the second leading cause of neurological disability after trauma among young adults worldwide (2).

1.1.1 Epidemiology

The prevalence of MS varies extensively across different populations and correlation with geographical latitude has been observed. High risk areas are found at higher latitudes, and in the Swedish population, which is one of the high risk Northern European populations, the prevalence has been estimated to 188.9/100,000 individuals, with a female to male ratio of 2.35:1 (3). Exceptions to the latitude gradient correlation are found in some isolated populations such as the Sami in Northern Scandinavia with a very low prevalence of MS (4) or the Sardinian population in southern Europe with a high prevalence (5). In the Swedish Multiple Sclerosis Registry the mean age at onset of disease was 33.7 years, as reported recently (6).

1.1.2 Clinical manifestations

Patients initially present with an attack of a neurological symptom or symptoms that can be of various kinds such as muscle weakness, numbness, paresthesia, optic neuritis or fatigue. A diagnosis of MS is defined based on a set of criteria which are still subject to revisions. According to the latest revision of the so called McDonald criteria, MS diagnosis is made if there have been more than two attacks with additional evidence of two or more central nervous system (CNS) white matter lesions visualized by magnetic resonance imaging (MRI). If one of these criteria is missing, additional criteria such as dissemination in time and space of visualized lesions and in some cases presence of oligoclonal bands (OCBs) or elevated IgG index in the cerebrospinal fluid (CSF) may be used (7). Patients who have experienced an attack but do not fulfill all criteria for MS diagnosis are classified as suffering from the clinically isolated syndrome (CIS).

Approximately 80% of patients diagnosed with MS have a relapsing-remitting MS (RRMS) disease course, characterized by bouts of diseases symptoms (relapses) followed by complete or partial recovery from symptoms (remissions). Relapse rate, symptoms and acquired disability varies to a high degree between patients. With time a majority of RRMS patients enter a progressive disease course termed secondary progressive MS (SPMS) in which the recovery phases are absent while there is a constant increase in disability. Twenty percent have a so called primary progressive MS (PPMS) course with an evident progression from disease onset (8). Disability in MS patients is measured according to the Expanded Disability Status Scale (EDSS), which ranges from 0.0 (no neurological symptoms) to 10.0 (death due to MS) (9).

OCBs in the CSF can be visualized by using isoelectric focusing on agarose gel followed by immunoblotting or immunofixation for IgG. The bands correspond to immunoglobulins produced by various plasma cell clones and can be found in the CSF of >90% of all patients with MS (10).

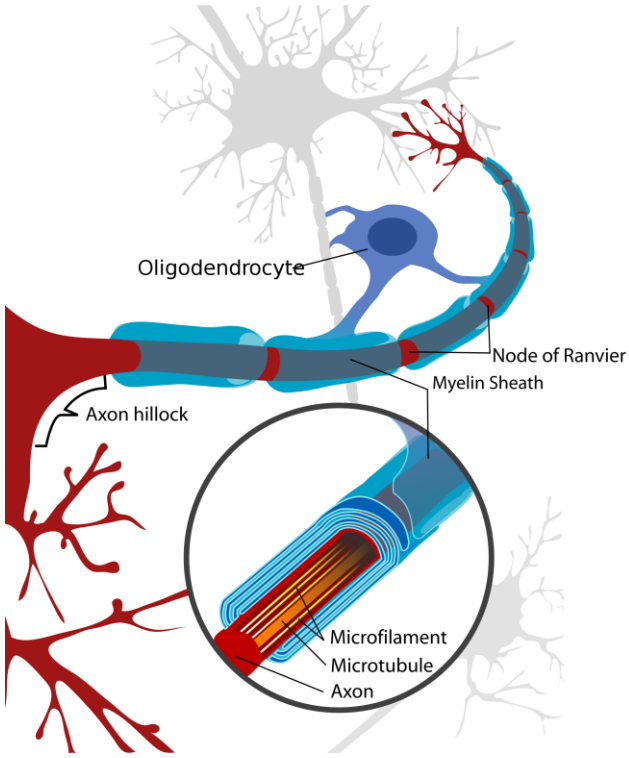

Figure 1, Schematic picture of a nerve axon with myelin sheats generated by an oligodendrocyte. Source: Wikimedia Commons, http://commons.wikimedia.org/wiki/Category:Images

1.1.3 Immunopathology

The lesions found in the CNS of MS patients arise due to immune attacks on myelin sheaths surrounding nerve axons (Figure 1). Disease symptoms are believed to be a consequence of compromised action potential conduction due to loss of the isolating myelin, and eventually due to axonal loss.

Under normal conditions, the CNS is isolated from the peripheral immune system through the tight blood brain barrier (BBB), however in MS this barrier breaks down, which facilitates immune cell trafficking into the CNS.

Lesions are most often found around blood vessels through which lymphocytes and macrophages infiltrate. In active lesions large numbers of myelin debris-containing macrophages have been observed (11). Clonally expanded $CD8^+$ T cells are most abundant cells in lesions, and $CD4^+$ T cells are present to a lesser degree (12). γδ T cells (13) as well as monocytes have been found, while B cells and plasma cells are less common (14). Based on conclusions from the common animal model for MS, experimental autoimmune encephalomyelitis (EAE), Th1 cells have previously been thought to drive the inflammation in MS (15), however convincing data also support a critical role of pro-inflammatory Th17 cells in initiation of MS (16, 17). A loss of suppressive activity of T_{regs} in response to autoreactive T cells has been found in patients with MS (18, 19).

The role of B cells in MS is a matter of current debate and investigation (20, 21). Successful clinical trials of B cell-targeting monoclonal antibodies for treatment of MS have underscored their importance (22, 23). Lymphoid follicles have been found in the meningeal areas in brains of SPMS patients (24), suggesting a persistence of potential antigens that drive humoral response (25).

It is obvious that MS is an inflammatory disease, however it is being debated whether the initiating event is a loss of tolerance in the periphery or whether it starts with neurodegeneration which leads to inflammation. In an autoimmune scenario, myelin-specific T cells that have not been eliminated by tolerance mechanisms become activated in the periphery and migrate to the CNS where they are reactivated and initiate inflammation. A possible mechanism for the peripheral activation of autoreactive cells could be molecular mimicry (26). As mentioned earlier, while currently available immunomodulatory treatments have proven to be effective in reducing relapse-rate and postponing progression in patients with RRMS, there is no treatment for patients with SPMS or PPMS. It has therefore been suggested that progressive MS is driven by different mechanisms than RRMS (8, 27, 28).

1.1.4 Treatments

There is no cure for MS, however currently available treatments that all target the immune system, can decrease symptoms and postpone or prevent progression in many, but far from all patients with RRMS. There is no effective treatment for patients with a progressive disease course.

IFN-β is the most commonly used treatment, proven to decrease relapse rate with approximately 30% (29, 30). The mechanism of the clinical effect of IFN-β is not completely understood although several have been proposed (31). The treatment response to IFN-β varies. Some patients are non-responders and some develop neutralizing antibodies against the drug. Another treatment with similar long-term efficacy to that of IFN-β is Glatiramer acetate, a four amino acid long peptide that can bind HLA class II molecules and modulate MS through a poorly understood mechanism (32). Natalizumab is a monoclonal antibody against the alpha-4 integrin subunit on the surface of lymphocytes. Blocking of this receptor for the VCAM-1 adhesion molecule impairs migration of lymphocytes through the BBB. Treatment with Natalizumab has showed a striking reduction of relapses and detectable lesions by MRI (33). However, progressive multifocal leukoencephalopathy (PML) is a rare but extremely serious side effect caused by an opportunistic JC virus infection of the CNS in absence of immunosurveillance (34). Other, less commonly used treatments include Fingolimod, the first oral MS, mitoxantrone and alemtuzumab, a lymphocyte-depleting monoclonal antibody. Clinical trials for several monoclonal antibodies for MS treatment have been performed, such as the B-cell targeting rituximab and ocrelizumab. These showed promising results but are not used as standard treatments for MS (22, 23).

1.2 ETIOLOGICAL STUDIES OF MULTIPLE SCLEROSIS

No single genetic variant or external factor has been identified as the cause of MS, however several factors that are associated with increased disease risk on a population level have been found. Not all of the risk factors are found in one individual with MS, and many healthy persons are exposed to the same risk factors. However in comparison with healthy persons, the frequencies of the risk factors are higher among persons with MS. This has led to the definition of MS as being a multifactorial, complex disease, meaning that it is likely caused by a complex interaction between environmental and genetic factors. Rothman et al have proposed a useful conceptual model for causes of multifactorial diseases. This thesis is to a large extent influenced by this so called “pie model” or “sufficient-component cause model”

(35), visualized in Figure 2. The important message of this model is that there may be many different combinations of risk factors that are necessary to act together for development of disease. Different individuals with the disease will have different combinations of risk factors, and each of those combinations is sufficient to cause the disease. In the following sections I will present some of the risk factors that have been identified for MS, with specific focus on the genetic risk factors.

Figure 2, Rothman's pie model. Each pie represents a sufficient cause of disease. Each slice in a pie represents a risk factor. In one individual, the combination of all the risk factors present in one sufficient cause is necessary for development of disease. Some of the risk factors overlap between different sufficient causes.

1.2.1 Environmental factors

Vitamin D/sunlight deficiency, cigarette smoking, Epstein-Barr virus (EBV) infection and high body mass index (BMI) are the environmental risk factors for MS that are backed up with most evidence.

Vitamin D levels and sunlight exposure have been studied due to the latitude gradient of MS prevalence. Low exposure to sunlight decreases the endogenous vitamin D production, which has been associated with increased MS risk (36). Interestingly, low sunlight exposure has been suggested to be a stronger independent risk factor than vitamin D levels (37).

A number of reports show an association of cigarette smoking (38, 39), EBV infection (anti-EBNA1 antibodies or mononucleosis) (40, 41) and BMI (42-44) with MS.

1.2.2 Genetics

Undoubtedly genetics play an important role in MS risk. Familial clustering of the disease is observed to a certain extent and in the Swedish population the proband-wise concordance rate in monozygotic twins has been estimated to 15.38 while it was 1.69 in dizygotic twins. The sibling recurrence risk (λ_s) has been estimated to be 7.1 (6). Moreover, as mentioned

earlier, the low prevalence of MS among isolated populations in certain high risk environments or high prevalence in lower risk environments, point to the significant influence of genetic factors (45-49).

1.2.2.1 Genetic variation and single nucleotide polymorphisms

Variation in the human genome, which leads to genetic differences between individuals and populations, can be of several different kinds. One example is copy number variations, including insertions, deletions, duplications or amplifications of genes, bases or stretches of DNA (50). The smallest variations in the genome are called single nucleotide polymorphisms (SNPs) and denote a one-base difference at a locus (Figure 4).

Individual 1 A G G **T** C A T T
Individual 2 A G G **A** C A T T

Figure 4, representation of a single nucleotide polymorphism (SNP) with different alleles in two individuals.

SNPs are very common and widespread across the genome. They can be found in non-coding sequences as well as in coding sequences. If the base change in a coding sequence alters the DNA codon in such a way it leads to an amino acid change in the encoded protein, the SNP is denoted as *non-synonymous*. Conversely, if the SNP does not alter the amino acid sequence, it is denoted as *synonymous.* Importantly, SNPs in non-coding sequences may alter loci that are involved in transcriptional regulation or epigenetic control.
A SNP that is common in a population does not by itself alter biological functions in a deleterious way. Variants that have more severe effects or seriously affect survival, will be very rare in the population.

1.2.2.2 Linkage disequilibrium

During each meiosis several crossing over events (recombinations) take place, thereby creating new haplotypes (combinations of alleles on a chromosome). Markers that are physically close to each other on a chromosome are more likely to be inherited together than more distant markers since the probability of recombination is lower the smaller the distance between two markers. *Linkage disequilibrium* (LD) is a measure of the number of recombinations that have taken place in a population. Other factors that are reflected in the LD include population history and selection (high LD between two markers might due to essential functions attributed to the presence of both markers, which are thus conserved on the same haplotype across generations).

The basis for the measurement of LD is the calculation of the disequilibrium coefficient (D),

$$D=P_{AB}-p_1q_1$$

where A and B are two different markers, P_{AB} is the frequency of their co-occurrence on the same haplotype, and p_1 and q_1 are the allele frequencies of the markers in the studied population.

Two frequently used measures of LD, D' and r^2, are derived from D. D' is normalized to the theoretical maximum of D which is dependent on the allele frequencies, and can have values between -1 and 1. In papers included in this thesis, I have used r^2 which is a measure of correlation between two markers in a given population. It can have values between 0 and 1 (where 1 means perfect LD), and is calculated as follows:

$$r^2=D/[P_A(1-P_A)][P_B(1-P_B)]$$

Importantly, due to LD structure in the genome, the association of a marker with disease might actually be due to the association with another marker in LD which is the true causal variant that contributes to pathogenesis. However, LD structure can also be a very useful tool for inference of alleles on a haplotype, called imputation (described in the Methodological Considerations section).

1.2.2.3 The HLA region

The strongest genetic risk factors in MS were already in the 1970's pinpointed to consist of alleles within the region of the major histocompatibility complex (MHC) on chromosome 6 (51-53). The MHC spans around 7.6 megabases (Mb) and contains 224 identified gene loci (54). A large number of these genes play central roles in the immune system, and of these the antigen presenting human leukocyte antigen (HLA) genes are the most important. MHC genes are subdivided into class I, II and III. Class I and II molecules are expressed on the cell surface and present peptides to T cells. Class I HLA genes are expressed on most nucleated cells in the human body and present endogenous peptides, which in most cases are self-peptides, but can also be intracellular pathogens such as viruses. Class II HLA genes are expressed on the surface of antigen-presenting cells (APCs) and present peptides from extracellular components that have been engulfed and degraded by the APCs.

The HLA genes are the most polymorphic genes known of in the human genome. According to the January 2014 IMGT/HLA Database report (http://www.ebi.ac.uk/ipd/imgt/hla/),

8,124 different alleles have been found among the HLA class I genes, being the *HLA-A, -B, -C, -E, -F, -G* protein coding genes and the *HLA-H, -J, -K, -L, -P, -T, -U, V, -W, -X* pseudogenes. For the HLA class II genes, being *HLA-DRA, -DRB1 to –DRB9, -DQA1, -DQB1, -DPA1, -DPB1, -DMA, -DMB, -DOA, -DOB* (some of which are pseudogenes), 2,409 different alleles have been found. The nomenclature of HLA alleles is periodically revised by the WHO Nomenclature Committee since new alleles are constantly being discovered. The current nomenclature system is illustrated in Figure 3.

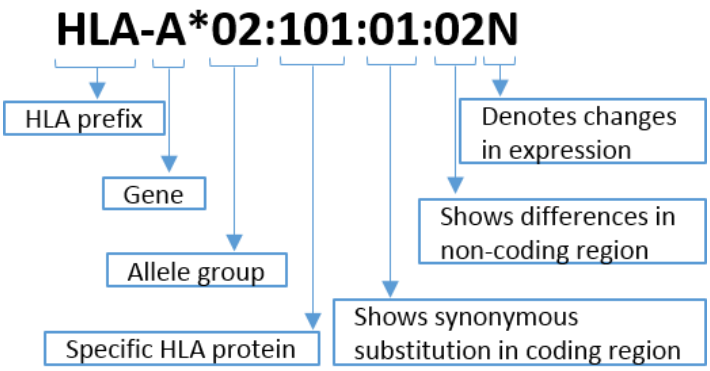

Figure 3, Nomenclature of HLA alleles. Figure adapted from http://hla.alleles.org/

The *HLA-DR2* haplotype (*DQB1*06:02-DQA1*01:02-DRB1*15:01-DRB5*01:01*) stands for the strongest genetic association with MS (55). It has been unclear which, if any, of the alleles in the haplotype, has an independent effect on MS risk, especially since the *DQB1*06:02* and *DRB1*15:01* alleles are in nearly complete LD in the Caucasian population. However, studies in the African American populations, as well as recent large scale studies corroborate an independent role of the *HLA-DRB1*15:01* allele (56-58). An independent strong protective association has been attributed to the *HLA-A*02* allele (59, 60), and has been confirmed (56, 57). The 2011 MS GWAS confirmed independent effects from the *HLA-DRB1*03:01* and *HLA-DRB1*13:03* alleles (56) and other alleles of the *HLA-DRB1* and *HLA-B* genes have been added the list of risk loci since (57).
The association with the *HLA* locus is a feature that MS has in common with many other complex inflammatory diseases, especially those that are defined as autoimmune, thus supporting the arguments for classification of MS as an autoimmune disease.

1.2.2.4 Genome-wide association studies

As it was early obvious that associated HLA alleles are not sufficient to explain the genetic risk for MS, both genome-wide linkage analyses and candidate gene association studies

have been undertaken. These were often inconclusive due to lack of power. Alongside the development of high throughput genotyping technologies, the genome-wide association studies (GWAS) were developed in order to screen for SNP associations with modest effects in complex phenotypes across thousands of samples, and the first GWAS was published in 2005 (61). The GWAS is a hypothesis-free approach in which SNPs that tag LD blocks across the genome are chosen to be present on an array, thereby covering the whole genome. Allele frequencies at all those SNPs are compared between cases and controls and statistically tested for disease association. Importantly, large-sized datasets of cases and controls as well as measures for controlling for population stratification are needed in order perform well-powered GWAS. The enormous number of association tests made in one GWAS makes it almost impossible to establish significant findings if common methods for multiple comparison corrections are applied. Based on a correction for the number of independent segments in the genome, as well as assumptions regarding the expected number of disease-associated loci and approximate numbers of cases and controls in a GWAS, a significance cutoff at $p<10^{-8}$ is applied in GWAS (62). The first MS GWAS was performed in 2007 and identified two susceptibility SNPs outside the HLA region, one mapping to the *IL7R* gene and another mapping to the *IL2RA* gene (63). After the most recent GWAS for MS a total of 57 non-*HLA* loci had been found, and four risk alleles in the *HLA* were confirmed (56) followed by the identification of three additional non-*HLA* loci through a meta-analysis (64). Most of these loci are located in, or close to, immune-related genes, which has led to viewing them as potential candidate genes for involvement in pathogenesis. An important aspect in association studies such as GWAS, is that only variants that are common in the studied population will be possible to detect if they are associated with disease. Moreover, GWAS chips are designed to only contain tag SNPs for common variants. Rare variants that contribute to disease may be present in subpopulations but will not be detected in association studies if the variants are rare enough.

1.2.2.5 The ImmunoChip Study

SNP arrays used for GWAS will only measure a subset of all the variants in the genome. Disease-associated SNPs found through these studies tag LD-blocks in which any variant could be causal in pathogenesis. ImmunoChip is a custom-made SNP microarray designed specifically to study loci with significant or suggestive evidence of association with any of 11 different autoimmune- or inflammatory diseases. One of the goals of the ImmunoChip was to fine-map regions where autoimmune- or inflammatory disease associated GWAS variants had been found, and another goal was to perform a deep replication of previous GWAS results (65). In Paper III and IV of this thesis I have used the 109 most associated

susceptibility variants found as a result of the MS-specific analysis which included 14,498 MS cases and 24,091 healthy controls (66).

1.2.2.6 Missing heritability and gene-gene interactions

As mentioned earlier, the sibling recurrence risk for MS has been estimated to be 7.1 according to the most recent and extensive study of the Swedish population (6). In the ImmunoChip study, a logistic regression model that included the 109 most significantly associated non-*HLA* susceptibility SNPs as well as the four *HLA* risk alleles established in the 2011 GWAS for MS (56), was used to estimate how much of the genetic variance can currently be explained by these variants. The null model used the relevant principal components and accounted for country strata. The conclusion from this analysis was that only 18% of the variance in the genetic analysis can be explained by the non-HLA SNPs and 27% if the HLA risk alleles are added to the analysis. In that analysis a λ_s of 6.3 was assumed (66). In other words, the increased risk for a sibling to an MS patient, who on average share 50% of the genetic setup, is 6.3 times that of non-related individuals, and only a fraction of that risk can be explained by the susceptibility variants we know of today. What about the remaining, unexplained fraction of the genetic risk? This is what we call the *missing heritability*.

There may still be many variants with small effect sizes or rare variants with large effects that remain to be identified in MS. Effects of epigenetics such as methylation or parent-of-origin effects (67, 68) may also form part of the explanation of the missing heritability problem found in studies of complex phenotypes in general. Another aspect that might contribute to the explanation of missing heritability, is a potentially increased risk due to interactions between genetic variants and environmental risk factors. In MS a few studies have shown interaction effects between *HLA* risk alleles and environmental risk factors such as smoking and exposure to EBV infection (40, 69). In Paper III of this thesis I have studied yet another possible explanation of missing heritability, which is the effect of gene-gene interactions.

There is some confusion regarding the use of terminology and definitions of genetic interactions. When the phenotypic expression of a gene is dependent on the presence (or absence) of one or several other genes, it is usually called epistasis (70). Sometimes the same term is used for merely statistical measures of interaction effects. The term gene-gene *interaction* might in some cases be thought of as a physical interaction between molecules in a biological pathway, however it is often used interchangeably with the term epistasis, i.e. also when the readout is a statistical measure. I have chosen to use the term gene-gene interaction or epistasis according to the definition that two loci interact if their joint effect

(in terms of disease risk in this case) is different than the sum or the product of their individual effects. Different can mean either more than, or less than the expected joint effect.

Gene-gene interactions can be studied by several different models, and in Paper III we have used two: a multiplicative model and an additive model. The multiplicative model assumes that the joint effect of two risk factors (in this case genetic variants x_1 and x_2), is expected to be the product of the separate effects attributed to each risk factor: $OR_{X1}xOR_{X2}$. Interaction according to the multiplicative model is defined as departure from multiplicativity of effects. In the additive model interaction is defined as departure from the expected additivity of two separate effects attributed to each genetic risk factor. Estimation of interaction on the multiplicative scale is a hypothetical measure since there is no estimation of the OR based on individuals who carry the supposedly interacting variants. ORs for each separate variant, as well as a third term which is the product of the two ORs, are estimated and fitted into a logistic regression model. This method can be useful in order to screen for potential interactions to test by comparing cases and controls. When estimating interaction on the additive scale, cases and controls carrying all possible combinations of two risk alleles studied, are compared to each other, and thus an OR for the co-occurrence of two alleles can be estimated. If that OR departs significantly from the additivity of separate risks, then these two alleles are supposed to belong to the same sufficient cause as defined by Rothman et al (35).

1.2.3 From genetic variation to function

1.2.3.1 Transcriptomics and eQTL analyses

Transcriptomics is defined as the study of the transcriptome, which is the total set of genes that have been transcribed to RNA and are present at the time of measurement in a given organism, tissue or cell. Transcription of genes is subject to an extremely stringent multilevel regulation which is tissue and cell specific, and it changes upon influences from different external stimuli. Changes in the transcriptome may also reflect pathological processes that are ongoing in cells or populations of cells. Since the advent of high throughput microarray technologies in the mid 90-ties (71) tens of thousands of transcriptome studies of different human tissues, physiological conditions as well as pathological conditions have been made. In studies of MS numerous expression profiling analyses have been performed, mainly in immune cells, e.g. comparing MS patients to healthy controls, or patients treated with IFN-β to untreated patients. However, differentially expressed genes in immune cells of MS cases and healthy controls are seldom found to overlap between different studies, as was

systematically reviewed in (72), which has led to the prevailing lack of pinpointed biological pathways in MS. As the transcriptome is highly dynamic, in contrast to the mostly static genome, common genetic variants that are associated with disease, are also common among healthy controls and variation in expression levels will therefore also exist among controls. All patients do not share the same susceptibility variants which leads to variation among cases, and many genes will be differentially expressed due to the inflammatory state in MS.

Expression quantitative trait loci (eQTLs) are features in the genome that influence the expression of genes. For instance, eQTLs can be a polymorphism or indel in a transcription factor binding site (TFBS) that will alter the binding of a transcription factor which is regulating a certain gene, and thereby affect the expression of that gene. eQTLs can act in *cis* which means that the expression of a nearby gene is affected, or in *trans* when more distant genes, usually on other chromosomes, are affected. The terms are used in a broad sense without set distance limits. *Cis* expression is more often studied compared to *tans* expression due to statistical and computational reasons. In whole-transcriptome studies of *trans*-acting eQTLs the number of SNP-gene expression comparisons will lead to a severe multiple testing penalty and the analysis requires large computational capacity. Progress is however made with regard to *trans*-eQTL analysis including a report supporting spacial clustering of *trans*-eQTLs near the MHC (73). It has been shown that eQTLs are tissue- and cell-specific to a large extent (74), although many eQTLs can also be replicated across tissues. As a logical consequence of the results obtained from GWAS of many common traits and diseases, a need for whole transcriptome eQTL analyses has become evident, and the currently available literature comprises several such studies (74-77). It has now become increasingly clear that regulatory variation such as eQTLs is extremely common in the human genome and that many known GWAS variants for different traits are eQTLs (78). In recent years RNA sequencing (RNA-seq) has emerged as a useful technique for transcriptome studies. This high throughput next generation sequencing technique with a higher sensitivity and possibilities to study transcription of genes that are not present on a microarray (79), was used to study *cis*-eQTLs in Paper IV in this thesis.

1.2.3.2 Biomarkers

A biomarker is defined as “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes or pharmacological responses to a therapeutic intervention” (80). The most widely used biomarker in MS is the MRI measurement. Other important measurements are the IgG index and analysis of OCBs

in the CSF. Several proteins found in the CSF, such as osteopontin (81, 82), matrix metalloproteinase-9 (82) and neurofilament (83, 84) have been studied as potential biomarkers in MS. In Paper II in this thesis I have used measurements of another potential biomarker under investigation, the CXCL13 chemokine, which is more abundant in the CSF of patients with inflammatory conditions in the CNS, such as MS, and that could potentially be used as a predictor of a severe disease course of MS (85, 86). In many cases, the objective of biomarker studies is precisely to find indicators that can be used directly in the clinic as an aid in diagnostics or prognosis prediction, however they may also be used as important indicators of pathophysiologic mechanisms of the disease.

1.2.3.3 Bioinformatics and integration of functional data

Following the completion of the human genome sequence, the ENCODE project is aiming to identify all functional elements in the genome (87). A plethora of functional annotations for genomic loci is now publicly available through genome browsers, such as http://genome.ucsc.edu/ or www.ensembl.org. The annotations are based on analyses performed in many different cell lines and primary cells. Of special interest are DNase I hypersensitive sites (DHSs), transcription factors found to bind the DNA using ChIP-seq, and histone marks specific for enhancers and promoters. Usage of this data can guide follow-up studies of disease-associated loci and identified eQTLs.

2 AIMS OF THE THESIS

The general aim of this thesis was to study genetic variants and their contribution to MS risk, through these specific aims:

I. To study whether the *IL21* locus, which has been associated with other autoimmune diseases, is also involved in risk of MS in a Swedish population

II. To study whether different MS-associated loci have a functional influence on levels of the potential disease severity biomarker CXCL13 in CSF of MS patients and thereby elucidate possible disease mechanisms

III. To study gene-gene interaction between MS-associated loci in order to search for causative pathways

IV. To analyze *cis*-regulation of gene expression by MS associated loci as a way to identify genes that are involved in pathogenesis

3 METHODOLOGICAL CONSIDERATIONS

3.1 CASES AND CONTROLS

DNA, RNA and CSF samples used in the studies presented in this thesis were included in one or several of the cohorts described below.

3.1.1 EIMS

Epidemiological Investigations in Multiple Sclerosis (EIMS) is an ongoing population based incident case-control study (39). Patients and controls have been recruited from all over Sweden. Cases fulfil the McDonald criteria (7). Controls are matched for age, sex, and residential area. DNA prepared from blood samples from this cohort were included in Paper I and III.

3.1.2 IMSE

This cohort includes MS patients who are treated with natalizumab and population based matched controls (age, sex, area of residence). The study subjects are recruited from all over Sweden. (88). In Paper I and III DNA from blood samples from this cohort was used.

3.1.3 BRAIN/STOP-MS

MS patients in these cohorts were recruited at the Karolinska University Hospitals at Solna and Huddinge and all fulfilled either the previously used Poser criteria (89) or McDonald criteria for MS (7, 90). The healthy controls are blood donors that visited any of three blood donation facilities in the Stockholm area in 2001 and 2004/2005 and were matched by ethnicity. The controls were not screened for MS. DNA from this cohort was used in Paper I and III.

3.1.4 GEMS

All cases in the Swedish MS registry who fulfil the McDonald criteria (7) and who have consented to participate in this study are included in GEMS (Genes and Environment in MS). The controls are population-based and matched for age, sex and residential area. (91). DNA from the GEMS cohort was used in Paper III.

3.1.5 STOP-MS

The STOP-MS cohort includes patients who have been recruited at the Karolinska University Hospital (Neurology Clinics of Solna and Huddinge districts), Stockholm, Sweden. The controls are not healthy matched controls, but patients with other neurological diseases, OND (non-inflammatory), or OND with signs of inflammation (iOND). In this thesis DNA, RNA from PBMCs and sorted cells, as well as cell-free CSF from the STOP-MS cohort were used. DNA samples were used in all the papers (in Paper I and III only MS cases from this cohort were included). In Paper II and IV RNA from PBMCs was used, and in Paper II CSF samples from this cohort were analysed.

The studies were approved by the Regional Ethical Board at Karolinska Institutet, Sweden. All subjects included in these studies had given their written or oral consent. All samples (separation of PBMCs, sorting of cells, DNA/RNA extractions) were prepared using standard procedures, described in detail in each of the papers in this thesis when applicable.

3.2 GENOTYPING

3.2.1 HLA genotyping

Classical typing of the *HLA-DRB1* and *HLA-A* genes was performed using sequence-specific amplification with the *Olerup* SSP® low resolution kits (Olerup SSP AB, Stockholm, Sweden) (92). Briefly, genomic DNA was mixed with a nucleotide-containing buffer and polymerase and added to wells pre-coated with sequence-specific primers. After a polymerase chain reaction (PCR) the products were run on agarose gel electrophoresis and visualized with GelRed (Biotium, Hayward, CA, USA) staining under UV light. The bands corresponding to specifically amplified HLA sequences were compared with a chart used for interpretation of the HLA alleles.

3.2.2 HLA imputation

Imputation of HLA types based on SNP genotypes from the *HLA* region typed on the ImmunoChip was used in Paper III and IV. The HLA*IMP:02 software (93) was used. In short, the SNPs from the dataset to be analyzed are aligned to a reference dataset (in this case the HapMap CEU panel), and subsequently the haplotypes are inferred and genotypes are imputed.

3.2.3 SNP genotyping

In Paper I and II a subset of the SNPs was genotyped using TaqMan SNP Genotyping Assays (Applied Biosystems, Foster City, CA, USA). A TaqMan assay includes probes with complementary bases for each allele of the SNP that is being analyzed. The probe is ligated to a VIC or FAM fluorophore, specified for each allele, which is coupled to a quencher that prevents emission of fluorescence. The assay also contains sequence-specific primers. A PCR was performed on genomic DNA samples and once primers and probe were bound and the sequence was amplified by the polymerase, the quencher was cleaved off and fluorescence was emitted. Allelic discrimination analysis in the samples was done by detection of fluorescent signals using the 7900HT Fast Real-Time PCR System and the SDS 2.4 software (both from Applied Biosystems, Foster City, CA, USA).

Another subset of SNPs in Paper I and II was analyzed with matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry (Sequenom Inc., San Diego, CA, USA) described in detail previously (94). This analysis was performed at the Mutation Analysis Facility, Karolinska Institutet.

The genotypes used in Paper III and IV were analyzed within the ImmunoChip study. DNA was hybridized to an Illumina Infinium SNP microarray (Illumina, San Diego, USA) with probes for 195,806 SNPs and 718 small indels. DNA samples from the cohorts used in this thesis were sent to the Wellcome Trust Sanger Institute and to the Miller School of Medicine, University Miami, where genotyping was performed. Quality control (QC) was done in Miami. Details on genotyping procedures and quality control can be found in (66).

3.3 EZYME-LINKED IMMUNOSORBENT ASSAY (ELISA)

For quantification of CXCL13 in CSF (Paper II) the commercially available Human CXCL13/BLC/BCA-1 Quantikine ELISA Kit (R&D Systems, Abingdon, UK) was used according to the manufacturer's instructions. Briefly, a CXCL13-specific monoclonal antibody was pre-coated onto a microtiter plate, upon which 50 µl undiluted cell-free CSF was added to the wells. After washing away unbound substances, an HRP-conjugated monoclonal antibody specific for CXCL13 was added. After washing, a substrate solution was added to the wells. As a consequence of the enzymatic reaction, color was developed in proportion to the amount of CXCL13 bound by the antibodies. The color intensities were measured on a spectrophotometer and CXCL13 was quantified.

3.4 GENE EXPRESSION ANALYSIS

3.4.1 Quantitative real-time PCR

cDNA was prepared using the iScript Reverse Transcription Kit (BioRad Laboratories, Hercules, CA, USA).

IQ SYBR Green Supermix (BioRad Laboratories, Hercules, CA, USA) was used in the real time qRT-PCR reactions performed in Paper II and the PCR was run on a BioRad iQ5 iCycler Detection System. SYBR Green binds to double-stranded DNA molecules and emits fluorescence which is quantified and is proportional to the amount of the product in the PCR reaction. Primer specificity was tested with melt curve analyses and PCR products were run on agarose gels in order to verify presence of single bands. GAPDH was used as an endogenous control. Expression was quantified with either the standard curve method (95, 96) or with the Delta-Delta Ct ($2^{-\Delta\Delta Ct}$) formula (97).

TaqMan Gene Expression Assays were used in the qRT-PCR reactions performed in Paper IV. The principle for TaqMan chemistry has been described under the SNP genotyping section. Fluorescence emitted by probes that have bound to amplified sequences is proportional to the amount of PCR products. The PCR was run on the CFX384 Real-Time PCR Detection System (BioRad, Hercules, CA, USA), and PCR reaction efficiencies, E (95),(98) were calculated using the BioRad CFX Manager System Software. 18S rRNA and HPRT1 were used as endogenous controls. Expression was quantified with a modified normalization method with two endogenous controls taken into account (99).

3.4.2 RNA sequencing

cDNA libraries for RNA-seq were prepared using the Illumina TruSeq kit (Illumina, San Diego, USA), according to the manufacturer's protocol.

Paired-end sequencing with 100 bp reads on an Illumina HiSeq 2000 machine (Illumina, San Diego, USA) was performed with an average sequence depth of 36 million reads per sample. All data processing and QC procedures have been described in detail in Paper IV. cDNA synthesis and RNA-seq was performed at the Science for Life Laboratory, Stockholm, Sweden.

After mapping of the obtained reads to the H. Sapiens reference genome (NCBI v.37, hg19), gene counts were computed (100) and normalized using the trimmed mean of M-values (TMM) methodology (101).

3.5 STATISTICAL ANALYSIS

3.5.1 Case-control association tests

In Paper I and III the - -assoc command in PLINK v.1.07 software (102) was used for calculation of odds ratios (ORs) based on chi-square (χ^2) tests.

3.5.2 Meta-analysis

The meta-analysis in Paper I was performed using a fixed-effect Mantel-Haenszel test with the meta.MH command in rmeta package in the free software R (103).

3.5.3 Quantitative trait association analysis

In Paper II a quantitative trait association test CXCL13 levels across genotypes (0,1,2) was performed with the - -assoc command in PLINK v.1.07 software (102). The command applies a Wald test to calculate p-values.

3.5.4 Interaction analyses

The gene-gene interaction analyses in Paper III was done using a modified and JAVA-coded version of the Gene-Environment and Gene-Gene Interaction Research Application (GEIRA) algorithm (104), which estimates interaction using two different methods. Both methods are based on calculation of ORs using logistic regression. For estimation of interactions on the multiplicative scale, a third term, the interaction term, is added to the logistic regression model and tested for significance. For estimation of interactions on the additive scale, the ORs are used to calculate the relative excess risk due to interaction (RERI), which in turn is used for calculation of the Attributable proportion due to interaction (AP).

3.5.5 Generalized linear model for eQTL analysis

In Paper IV a generalized linear model was used to estimate correlation between genotypes and gene expression levels, assuming an additive genetic model. Genotypes and covariates were included as independent variables and the expression data, the dependent variable, was assumed to follow a negative binomial distribution.

3.5.6 Multiple comparisons, permutations and bootstrapping

Several different methods were applied in Paper II, III and IV in order to control our statistical measures.

In Paper II the false discovery rate (FDR) was controlled for by calculation of Benjamini-Hochberg adjusted p-values using the multtest package in R (103).

In paper III 1,000 permutations were applied to the data, and the permuted p-values were corrected for multiple testing. Corrected permuted p-values correspond to the proportion of permuted p-values over all markers that were equal or more extreme than the estimated p-value. A significance cut-off at $p<0.05$ was used.

Permutations were also applied to the data in Paper IV. Moreover the strength of the correlation estimates was tested using bootstrapping (resampling over covariate-stratified subgroups), and FDR was estimated with a method designed for non-parametric data, described in (105).

4 RESULTS AND DISCUSSION

4.1 PAPER I

There were two main reasons that lead us to investigate the *IL21* locus and nearby genomic region with regard to association with MS: Firstly, our group had previously found a rat EAE QTL containing the *Il21r* and then found a haplotype in the *IL21R* region that was associated with MS (106), which directed our interest towards the IL-21 pathway. Secondly, while polymorphisms in the *IL21R* and *IL21* regions are risk loci in systemic lupus erythematosous (SLE) (107, 108) they had also been found to associate with high serum levels of IL-21 in MS patients, which was predictive of development of a secondary autoimmune disorder post treatment with a lymphocyte-depleting antibody. However, there was also a negative report from a study in a rather modestly sized Spanish cohort, showing lack of association with MS in this region (109).

Results from Paper I

We thus investigated this by genotyping 12 SNPs in a region comprising the *KIAA1109*, *ADAD1*, *IL2* and *IL21* genes, in 2090 Swedish MS cases and 1732 healthy controls. Seven of the SNPs were tagging the *IL21* gene, three additional SNPs were chosen because they were associated with IL-21 serum levels, and two additional SNPs were added in order to fill genomic gaps. With this number of subjects we had 93% power to detects an association of the same effect size (OR=1.6) that was found for one of the *IL21* SNPs in SLE (107), however for detection of an OR=1.2, we only had 40% power. We performed an association analysis and did not find significant associations for any of the SNPs with MS, as shown in Table 1 in Paper I. We also evaluated whether *IL21* polymorphisms are involved in epistatic effects that would increase risk for MS only together with the MS associated *IL21R* polymorphism or the *HLA-DRB1*15* variant. We did not find any signs of epistasis between these loci. Finally, we performed a meta-analysis in which we used allele frequency data for one SNP in the published Spanish study (109), that overlapped with one of the 12 SNPs that we had typed in the *IL21* region. This pooled analysis including 1983 cases and 1682 controls from Sweden, and 768 cases and 929 controls from Spain, confirmed the lack of association found in each of the separate studies (Figure 5).

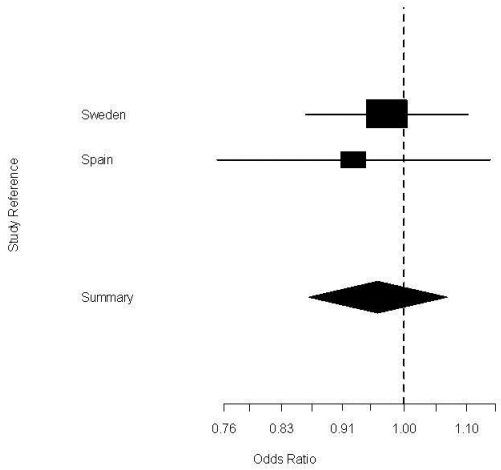

Figure 5, Meta analysis of rs6822844 in the IL21 region, genotyped in 1983 cases and 1682 controls from Sweden, as well as 768 cases and 929 controls from Spain. The pooled analysis confirms the lack of association with MS.

Discussion on Paper I

This paper reports a so called "negative result". We had a hypothesis, we tested it and could not prove it. We also compared our analysis with another, smaller one done before us, and confirmed the negative result: An association of variants in the *IL21* region with MS could not be supported by our data. By reporting this lack of association, we want to contribute to the knowledge about genetic risk factors in MS and prevent other researchers from spending time and money on trying to find associations with MS in the *IL21* region, unless they use a rather distant population. Our study had low power to detect a potential small or moderate effect in the *IL21* region, however later data from the MS GWAS and the ImmunoChip study in MS confirmed that *IL21* does not harbor any MS risk locus in Caucasians (56, 66). Moreover, none of these studies found genome-wide significant associations in the *IL21R* region, a region that our group had previously reported a nominal association to (106), thus weakening the probability of the *IL21R* being a candidate gene in MS. However, MS is a highly heterogeneous disease and we cannot exclude that the *IL21* region might be involved in risk in a certain subgroup of patients.

4.2 PAPER II

The B cell-attracting chemokine CXCL13 has been investigated by quite a few researchers as a potential CSF biomarker for MS and other inflammatory conditions in the CNS (85, 86, 110-113). Although it cannot be used to distinguish between different neuroinflammatory

diseases (111), CXCL13 has been shown to have a predictive potential for prognosis in CIS and MS, where increased CSF concentration was found more often in CIS patients who converted to RRMS, and in MS patients with higher disease activity (85, 86).

Our aim was to investigate whether some of the genetic variants that were known or suggested to be associated with MS at the time of starting the investigation (before the latest GWAS from 2011 (56)), also associate with levels of CXCL13 in the CSF. Why would that be of interest? From a clinical point of view, such information could reinforce a set of markers that could potentially be useful in prediction of prognosis for the patient. From the basic researcher's point of view, combination of genotypic and phenotypic data is useful for pinpointing biological pathways in which CXCL13 and the genetic risk variants for MS are involved, and thereby guide further investigations of disease mechanisms.

Results from Paper II

We genotyped *HLA-DRB1* and *HLA-A*, as well as 23 non-*HLA* SNPs, and measured the concentration of CXCL13 in the CSF of 663 individuals diagnosed with MS, CIS, iOND or OND. We then used the genotype data for correlation with the CXCL13 levels. We did not find significant correlations in the set of patients with iOND and OND (total number=253), while five markers (Table 1) showed a significant genotype – CXCL13 level correlation in patients with MS and CIS (total number=410). When all disease groups were pooled together, one additional SNP was significant for genotype correlation with CXCL13 levels (Table 1). Thus, in summary, we found that carrying MS risk genotypes at SNPs in the gene region of (or close to) *RGS1, IRF5* and *OLIG3/TNFAIP3* as well as carrying the *HLA-DRB1*15* allele was significantly associated with higher levels of CXCL13 in the CSF. These effects were pronounced in patients with MS or CIS, and for some of the markers, the significance of the association was increased by adding OND and iOND to the analysis. As presented in Table 1, and as expected since the variants are associated with MS, genotype frequencies differed slightly in MS and CIS patients as compared to OND and iOND. This, together with the fact that the group of patients with OND and iOND was smaller than the group with MS and CIS, and have lower levels of CXCL13 (Figure 1, Paper II), might explain why significant associations between the variants and CXCL13 were not found in the smaller patient group.

		All diagnoses				OND and iOND				MS and CIS			
		Genotype frequencies			CXCL13 correlation	Genotype frequencies			CXCL13 correlation	Genotype frequencies			CXCL13 correlation
SNP	Closest gene(s)	11	12	22	ADJ. P	11	12	22	ADJ. P	11	12	22	ADJ. P
rs2760524	*RGS1*	0.02	0.27	0.71	0.02	0.02	0.32	0.65	0.80	0.01	0.24	0.75	0.14
rs9321619	*OLIG3/ TNFAIP3*	0.24	0.47	0.29	0.06	0.22	0.46	0.32	0.80	0.25	0.48	0.27	0.04
DRB1*15	*HLA-DRB1*	0.07	0.40	0.53	0.02	0.02	0.29	0.69	0.86	0.10	0.45	0.45	0.04
rs4728142	*IRF5*	0.22	0.49	0.29	0.02	0.24	0.48	0.28	0.80	0.21	0.49	0.29	0.02
rs3807306	*IRF5*	0.22	0.50	0.29	0.02	0.23	0.47	0.30	0.81	0.21	0.52	0.28	0.01

Table1, Genotype frequencies for the variants that were associated with CXCL13 levels in the different patients groups. ADJ. P: adjusted p-value

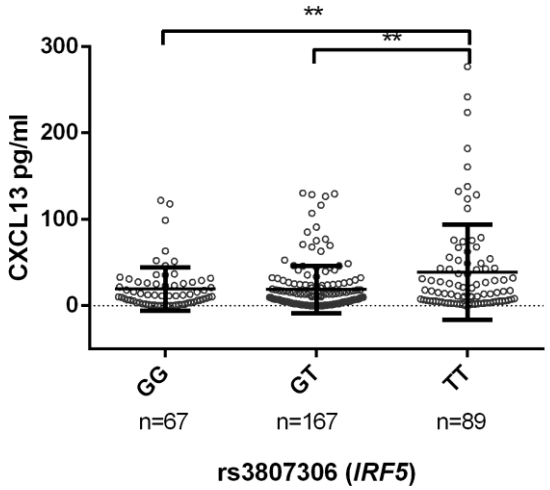

Figure 6, Representative graph showing association of CXCL13 levels in the CSF with genotypes at the rs3807306 locus. Mean with SD is shown. Differences between groups were tested with the Student's t-test. This analysis was done in patients with MS and CIS.

Discussion on Paper II

Our results show an association between genetic susceptibility variants in MS and increased CXCL13 levels in the CSF, which in MS patients correlates with the severity of the disease course. We do not know whether the associated SNPs are causal in regulating CXCL13 or whether they are in LD with causal SNPs or haplotypes. We speculated about potential functional connections between the genes located close to CXCL13-associated variants and

the chemokine itself. Two studies have suggested a link between *HLA-DRB1*15* positivity and a more severe disease course of MS (114, 115), which is what could connect this allele to the increased levels of CXCL13 (86). *RGS1* encodes a protein that attenuates signaling through CXCR5, the receptor of CXCL13 (116, 117). IRF5 is a transcription factor that binds to interferon-sensitive response elements (ISREs) close to many genes involved in proinflammatory responses. Putative ISREs are found in proximity of the CXCL13 gene (http://genome.ucsc.edu/), however experimental data confirming a direct regulation of CXCL13 by IRF5 is to our knowledge lacking. The SNP rs9321619 maps to an intergenic region. The closest genes are *OLIG3,* involved in nervous system development (118) and *TNFAIP3* (also called A20) which encodes a negative regulator of NFκB-mediated proinflammatory response of TNF (119) and mice lacking A20 in dendritic cells develop systemic autoimmunity (120). In view of more recent studies, it should be said that the *RGS1, OLIG3/TNFAIP3* and *IRF5* regions were all included in the ImmunoChip analysis (66) and two different SNPs in the *RGS1* and *OLIG3/TNFAIP3* regions were more significantly associated with MS than rs2760524 and rs9321619, respectively. LD between rs2760524 and the most associated ImmunoChip SNP in the region was however very high (r^2=0.959). rs9321619 was not typed on the ImmunoChip and the intergenic region where it is located had very few SNPs typed on the chip. Using genotype data from the 1000 Genomes, we found a SNP that was typed on the ImmunoChip and is in high LD (r^2=0.953) with rs9321619, while it is in very low LD (r^2=0.007) with the most associated SNP on the ImmunoChip, mapping nearly 100 kb away from rs9321619. The *IRF5* genetic region has not reached genome-wide significance for association with MS, only nominal significance in smaller cohorts (121). However, in the ImmunoChip study several markers in the *IRF5* region reached p-values of 10^{-4}. *HLA-DRB1*15* certainly keeps being the most associated variant in MS, with highly probable, yet largely unexplored functional implications in pathogenesis. In Paper II we discussed the functions of the genes that are found closest to the SNPs associated with CXCL13 levels, however with the consciousness of the possibility that these genes are perhaps not affected by the SNPs. In the eQTL analysis that we have described in Paper IV, none of the most disease-associated SNPs in the *RGS1* and *OLIG3/TNFAIP3* regions affected expression of any of these genes when analyzed in PBMCs. The *IRF5* region was not included in the eQTL analysis (Paper IV) since it was not associated with MS with genome-wide significance. The *HLA-DRB1*15* variant influenced expression of different genes in the *HLA* region (Paper IV). We cannot exclude though, that eQTL effects including these SNPs and genes could be found in other cell types which are not present, or very sparse among PBMCs. Moreover, it has not been studied whether other types of regulation such as splicing affects any of these genes through the SNPs.

CXCL13 found in the CSF is for sure a marker of generalized inflammation in the CNS. The associations between these genetic risk variants in MS and increased levels of CXCL13 still need to be replicated. What remains to be investigated if they become established, is whether these variants predispose to more inflammation, or less controlled inflammation, which in turn becomes both a risk factor for disease, and a risk factor for a more severe disease.

4.3 PAPER III

Since only a fraction of the heredity of MS can be explained by all the single genetic risk factors that have been identified to date, it is plausible that a part of this missing heritability might be explained by effects that are due to interactions between genes. In Paper III we aimed to evaluate gene-gene interactions between the currently known MS risk variants in a cohort consisting of 2081 cases and 2166 healthy controls of Swedish or Scandinavian origin. We used genotypes for the 109 SNPs that were typed within the ImmunoChip Project and that showed the most significant independent associations with MS (66), and imputed the established MS-associated *HLA-DRB1*15:01, HLA-DRB1*03:01, HLA-DRB1*13:03* and *HLA-A*02:01* alleles (56). Using this genotype data we tested gene-gene interactions with two different methods: estimation of the product term in a logistic regression model which measures interaction as departure from multiplicativity of the effects of two risk variants, as well as estimation of the attributable proportion due to interaction (AP) which measures interaction as departure from the additivity of effects.

Results from Paper III

We found significant interactions on the additive scale between presence of *HLA-DRB1*15:01* and absence of the protective *HLA-A*02:01*, presence of *HLA-DRB1*15:01* and the rs6677309 risk allele (A) and presence of *HLA-DRB1*03:01* and the rs7196953 risk allele (A). No interactions on the multiplicative scale remained significant after correction for multiple testing.

Genetic variants	AP	95% CI	p-value	permuted and corrected p-value
HLA-DRB1*15:01+ HLA*02:01 -	0.55	0.28-0.59	$1.69*10^{-8}$	0.011
HLA-DRB1*15:01+ rs6677309_A +	0.44	0.35-0.75	$4.76*10^{-8}$	0.013
HLA-DRB1*03:01+ rs7196953_A +	0.71	0.45-0.97	$1.01*10^{-7}$	0.018

Table 2, Significant interactions on the additive scale between pairs of genetic variants. The estimate AP denotes the Attributable Proportion due to interaction

Discussion on Paper III

In all three cases of significant interaction on the additive scale that we report here, a risk variant of *HLA-DBR1* is involved in the interaction, in two cases it is the *HLA-DRB1*15:01* allele and in one case it is the *HLA-DRB1*03:01* allele. It is well known that *HLA* alleles have the strongest effects on MS risk (OR=3.1 and 1.26 respectively, compared to OR<1.20 for most non-*HLA* SNPs). If we hypothesize that MS risk loci interact in causing the disease, and that *HLA* is involved in those interactions, then the probability to detect interactions with *HLA* alleles is higher compared to the probability of finding interactions among non-*HLA* loci with lower ORs. There is no existing method for power estimation in interaction studies using departure from additivity, however our results indicate that we do not have enough power to detect significant interactions between MS risk loci outside the *HLA* region.

Hedström et al have already reported an interaction between presence of *HLA-DRB1*15:01* and absence of *HLA-A*02:01* in a smaller subset of our case-control cohort, and they only found significant interaction among smokers (69). We again show this interaction with more cases and controls included, however we lacked data on smoking habits for a major part of the study subjects in our cohort, and therefore we have not stratified for smoking in our analysis. Thus, it remains to be investigated in an extended cohort using information on smoking habits, whether the interaction can be independent of smoking or not. How the *HLA-A*02:01* allele confers protection from MS is not known, however studies performed in transgenic mice found that *HLA-A*02:01* carriers had a dramatic reduction of autoimmune response compared to mice carrying the *HLA-A*03:01* allele (122) which, even though not established as such, is suggested as a potential risk allele in MS (59, 123). Lacking *HLA-A*02:01* might thus "interact" with *HLA-DRB1*15:01* through a lack of reduced immune response, and/or by effects from risk alleles in the *HLA-A* that would not exert effects in presence of *HLA-A*02:01.*

We identify interaction between rs6677309 and *HLA-DRB1*15:01* such that the risk associated with having risk alleles at both loci is higher than expected (Figure 2). In fact most of the *HLA-DRB1*15:01* risk is observed in the presence of the rs6677309 risk allele. The rs6677309 SNP, has been found to affect expression of the *CD58* gene (124, 125) which encodes a molecule that is involved in antigen presentation (126, 127), a function that it has in common with the *HLA-DRB1* gene that is encoding the beta chain of the HLA class II molecule. Expression of *HLA-DRB1* is influenced by the *HLA-DRB1*15:01* allele, as shown by us (Paper IV) and others (125, 128). Conversely, we did not observe an association with rs6677309 genotypes and *CD58* expression in Paper IV when analyzed in PBMCs. Even though our result has to be interpreted with caution due to the small number of individuals in the reference group (negative for both risk factors) used in the interaction analysis, the

reported functional consequences encourage experimental tests of the interaction between *HLA-DRB1*15:01* and rs6677309 A allele (Figure 7).

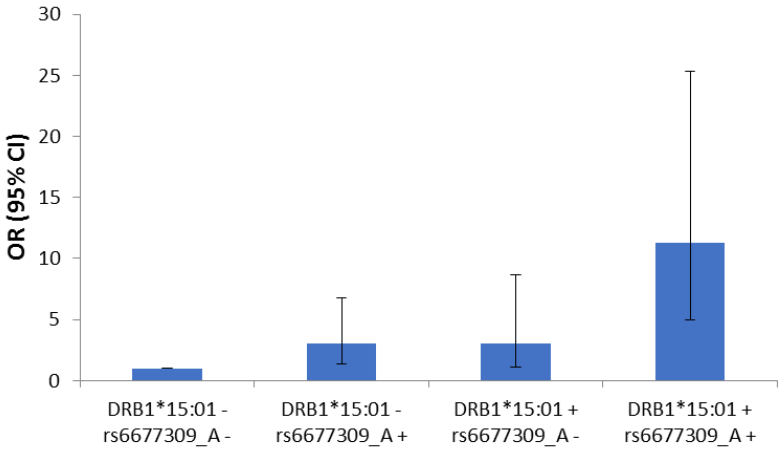

*Figure 7, Effects (ORs) of presence of the single variants HLA-DRB1*15:01 and rs6677309_A, and their interaction.*

In contrast to *HLA-DRB1*15:01* there are, to our knowledge, no published reports regarding the influence of the *HLA-DRB1*03:01* allele on expression of *HLA* genes, however in Paper IV we found that several *HLA* genes were affected both in PBMCs and in LCLs, and interestingly the effects were consistently opposite to those of the *HLA-DRB1*15:01* allele. We observe an interaction of the *HLA-DRB1*03:01* with the risk allele of rs7196953. No functional studies regarding rs7196953 have been reported. It is located in between the *MAF* gene and the gene and the *DYNLRB2* gene and does not map to any DHS that can be found in the public data provided by the ENCODE project (87).

Our goal is to understand more of the role of genetics and biological pathways involved in pathogenesis with the help of gene-gene interaction studies, however as I pointed out in the beginning of this discussion, these analyses probably require much more statistical power. Moreover, variants with lower effects, not reaching genome-wide significance or even any variant in the genome (not found in association studies) could be involved in interactions that lead to disease. Since we cannot calculate power, it is not known how many cases and controls that are needed to screen for such interactions, only that it has to be more than in the studies performed to date. Another major limitation in the screening for interacting loci, is the current technology and computational resources which make large numbers of statistical tests with the necessary number of permutations or resamplings rather infeasible.

One possible strategy is to allow less stringent significance thresholds and choose sets of potentially interacting variants with reported functions, and turn to experimental validations.

When it comes to studying interactions on a multiplicative scale, we are actually studying a hypothetical interaction, since we do not measure whether both risk alleles are carried together more often in those who are cases compared to those who are controls. We only measure the OR for each separate variant, and then measure whether the slope of the regression curve increases dramatically (more than expected) when the two ORs are multiplied with each other. If that is the case one can use the case-control cohort to test whether the simultaneous presence of the two risk factors is significantly more frequent among cases. We did not find interactions on the multiplicative scale which held for multiple testing corrections, and have therefore not investigated them further in the present study. It would however be interesting to explore the potential multiplicative interactions in the future, using an extended case-control cohort.

4.4 PAPER IV

To our knowledge there is still no published study in which the currently most up-to-date list of MS-associated loci have been evaluated as potential eQTLs. We hypothesized that a fraction of these risk loci are eQTLs for one or several genes, given that eQTLs are very common in the human genome (78) and most of the risk loci are not coding variants, and given that dysregulation of gene expression might underlie disease development. As described in the introduction, eQTLs can act in *trans* or *cis*, and we decided to study *cis* regulation within the limits of an 800 kb genomic window centered around each associated locus. This restriction reduces the probability of observing false positive associations compared to a case where expression of all the genes in the genome would be tested for both *trans* and *cis* regulation by the MS risk loci. Moreover, based on previously reported observations, most *cis*-eQTLs or DHSs were found approximately at 400 kb distance from the affected gene (129, 130).

Results from Paper IV

We performed RNA-seq in PBMCs from 183 individuals, diagnosed with MS, clinically isolated syndrome or other neurological diseases. We used genotypes from the ImmunoChip Project for 109 SNPs with the strongest independent associations with MS, and HLA-types that were imputed using ImmunoChip data. In the eQTL analysis we tested for association of the 109 SNPs and four MS risk alleles in the *HLA* region, with expression levels

of the genes in the established genomic window. By using a regression model, we could correct for relevant covariates such as gender, diagnosis and batch of preparation in the analysis. We used a set of criteria to choose results that we considered to be significant and that we wanted to validate. We chose 20 non-*HLA* loci that influenced the expression of in a total of 28 genes and three *HLA* variants that influenced the expression of seven different genes, however five of them were influenced by two different alleles.

SNP	Closest gene(s)	Genotype-gene expression association	Significant in cell type(s)
rs11587876	*DDAH1*	*WDR63*	PBMCs
rs3748817	*MMEL1*	*MMEL1*	PBMCs
rs7595717	*CNRIP1 (dist=40294), PLEK (dist=4845)*	*PLEK*	PBMCs, $CD4^+$
rs842639	*FLJ16341*	*AHSA2*	PBMCs
rs9989735	*SP140*	*SLC16A14*	PBMCs
rs1920296	*IQCB1*	*IQCB1*	PBMCs, LCLs
rs7665090	*NFKB1 (dist=13144), MANBA (dist=1040)*	*KRT8P46*	PBMCs
		RP11-10L12.6.1	PBMCs, LCLs
		MANBA	PBMCs, LCLs
rs71624119	*ANKRD55*	*ANKRD55*	PBMCs
rs11154801	*AHI1*	*AHI1*	PBMCs, LCLs
rs941816	*PXT1*	*ETV7*	PBMCs
		RAB44	PBMCs
rs1021156	*PKIA (dist=58302), ZC2HC1A (dist=2478)*	*FAM164A (ZC2HC1A)*	PBMCs, LCLs
rs694739	*PRDX5 (dist=7938), CCDC88B (dist=10457)*	*AP003774.1.1*	PBMCs, LCLs
rs11052877	*CD69*	*RP11-726G1.1.1*	PBMCs
		CLECL1	PBMCs, LCLs
rs12296430	*LTBR (dist=2768), CD27-AS1 (dist=44667)*	*ZNF384*	PBMCs, LCLs
rs201202118*	*TSFM*	*METTL21B*	PBMCs
		XRCC6BP1	PBMCs
		AVIL	PBMCs
rs1886700	*CDH3*	*CDH1*	PBMCs
rs4780346	*CLEC16A (dist=12760), SOCS1 (dist=59468)*	*RMI2*	PBMCs
rs12946510	*GRB7 (dist=8839), IKZF3 (dist=1591)*	*GSDMB*	PBMCs
		ORMDL3	PBMCs, LCLs, $CD19^+$
rs4794058	*MRPL45P2 (dist=27112), NPEPPS (dist=11346)*	*TBKBP1*	PBMCs, LCLs
		MRPL45P2	PBMCs, LCLs
rs470119	*TYMP*	*CPT1B*	PBMCs

Table 3, List of the most significant non-HLA SNP-gene expression associations presented in Paper IV

HLA variant	Genotype-expression association	Significant in cell type(s)
*HLA-DRB1*15:01 (presence)*	*HLA-DQB1*	PBMCs, LCLs
	HLA-DRB5	PBMCs, LCLs
	HLA-DRB1	PBMCs, LCLs
	HLA-DQB1-AS1	PBMCs, LCLs
	HLA-DQA1	PBMCs
*HLA-DRB1*03:01 (presence)*	*HLA-DQB1*	PBMCs, LCLs
	HLA-DRB5	PBMCs, LCLs
	HLA-DRB1	PBMCs, LCLs
	HLA-DQB1-AS1	PBMCs, LCLs
	HLA-DQA1	PBMCs
*HLA-A*02:01 (absence)*	*HCG4*	PBMCs, LCLs
	HLA-J	PBMCs, LCLs

Table 4, List of the most significant HLA variant-gene expression associations presented in Paper IV.

In the second phase of the project, we tested these potential eQTLs in an independent public dataset consisting of genotype and RNA-seq data from lymphoblastic cell lines from 232 individuals. We replicated 9 significant non-*HLA* eQTLs which affected the expression of in total 11 genes, as well as the three *HLA* variants and six genes influenced by them. In addition, we used an additional patient cohort (n= 59) to test three eQTLs in sorted PBMCs. Due to low minor allele frequencies for many of the SNPs, or very low expression levels, we could only test three of the SNPs in sorted PBMCs, where we were limited by the low number of samples and detection limits of qRT-PCR, the method we used for expression quantification in sorted PBMCs. We replicated two of the three eQTLs tested in sorted PBMCs. A summary of all significant is found in Table 3 and 4.

Discussion on Paper IV

In this paper we have used a hypothesis-free approach in the sense that we have included all genes present in the 800 kb windows in our eQTL analysis.

We performed our initial analysis in primary immune cells, which gave us the advantage of avoiding potentially cell-line specific effects, while the disadvantage lies in that PBMCs contain a mix of different cell types. Using cells from MS patients as well as patients with CIS and OND gave us the possibility to compare effects in MS patients with non-MS, while we had to correct for the different diagnosis groups in the regression analysis, thus reducing power. We could however conclude that most of the eQTLs observed in our study were independent of disease state.

In discussing the results from Paper IV, we must necessarily make some statistical considerations. Our data was influenced by several factors which we were aware of and intended to control for. With the help of a principal component analysis (PCA), we identified three factors among variables that we had data for, which significantly influenced variance in our data: batch of cDNA library preparation, gender of the patient and age at sampling. We corrected for these factors in the regression analysis by using them as covariates. In addition, we included disease state and interferon treatment as covariates, due to their likely influence on gene expression, especially of immune related genes. As commonly observed in gene expression studies, there was a widespread presence of outliers in the expression data for each gene. The outliers did not correspond to the same individuals across all genes. We used permutations in order to avoid false positive results that are in fact driven by outliers. Also because of the outliers, we did not know how robust our correlation estimates were and therefore we applied bootstrapping. Finally, in order to account for multiple testing, we applied a non-parametric FDR, which was based on the correlation estimates. Based on the FDR analysis, very few associations would be accepted at an FDR<0.25, on the other hand the lower FDRs more often correlated with high beta estimates. The criteria which we set for choosing associations to validate, prioritized a low permuted p-value over a low FDR or a high beta estimate. We assumed that small but robust effects (low beta values with low corresponding p-values) were likely to be true, even though they did not correspond to an acceptable FDR. However, to be able to evaluate whether strong effects (high beta estimates and low FDRs) would be replicated, we included some associations based on the latter criteria in the validation analysis. A detailed description of the validation criteria is found in Paper III. Most of the replicated SNP/HLA-variant gene expression associations corresponded to high FDRs (>0.25), low permuted p-values and varying beta estimates, thus confirming our assumptions, however without discarding the potential significance of high effects. We can also draw the conclusion that a thorough method for calculation of the FDR in this type of data is needed.

Many eQTLs are cell type specific or have more pronounced effects in specific tissues (74) and we believe that the lack of replication in LCLs of some eQTLs that were significantly associated in PBMCs reflect this cell specificity, which is also supported by one of the eQTLs, rs7595717 which was significantly associated with expression of *PLEK* in PBMCs and $CD4^+$ cells, but not in LCLs, $CD8^+$, $CD19^+$ or $CD14^+$ cells. We cannot exclude however, that the eQTL was active in any other of the sorted PBMC cell types, since the power in that analysis was very limited.

The genes that we have found in this study are involved in a wide range of different biological processes. Some of them have characterized functions in immune cells or

processes, such as *CLECL1* (131), *TBKBP1* (132, 133) and *ORMDL3* (134). Several genes are pseudogenes or genes with unknown functions, while other genes, such as *AHI1 (135, 136)* and *MMEL1* (137-139), seem to have functions described in different tissues and processes, including the CNS.

Expression of different *HLA* genes in relation to MS risk alleles require further investigation and replications before conclusions can be drawn, due to a partial lack of agreements between previous studies (refs).

Our findings show that several of the genetic variants associated with MS, affect the expression of genes that are potentially involved in the pathogenetic mechanisms. These results warrant further functional studies of the eQTL-regulated genes and the involved biological pathways, with focus on specific cell types.

4.5 SUMMARY OF FINDINGS

- We have shown that the *IL21* locus in not a major risk locus in MS in the Swedish population.

- We have found four MS-associated genetic variants that potentially influence pathways that lead to increased levels of the CXCL13 chemokine in CSF.

- We report three pairs of genetic variants that show interaction measured as departure from additivity of individual variant effects.

- We have identified 20 non-*HLA* MS risk loci and three *HLA* variants that in PBMCs show eQTL effects influencing expression of 28 non-*HLA* genes and seven *HLA* genes, respectively. We show that seven of the non-HLA eQTLs were also active in LCLs, one was active in $CD19^+$, and another in $CD4^+$ cells.

5 CONCLUDING REMARKS

In this thesis I have focused on the study of genetic loci that are associated with an increased risk of MS. When I started out my thesis work, less than 10 non-*HLA* risk loci in MS had been established (63, 140-142), and in screens for association with disease, the candidate gene approach was more widely used than today. Throughout these years new loci have been added to the list of risk variants, several associated regions have been fine-mapped and we now have increased possibilities for functional investigations of these loci. The studies included in this thesis reflect this development to some degree.

Throughout my work I have been searching for clues about which biological pathways to study in order to increase the understanding of MS disease mechanisms. For this I have used different strategies: candidate gene approach (Paper I), using biomarker data together with genetic associations to track back to potential disease mechanisms (Paper II), gene-gene interaction analysis (Paper III) and finally eQTL analysis (Paper IV).
In what way has this thesis contributed to the field of genetic research in MS? Hopefully, Paper I has helped to encourage efforts to search for other major effect candidate genes than *IL21*. Paper II-IV have resulted in interesting findings that prompt replication and/or functional investigations. We have found genetic associations with CSF levels of the CXCL13 chemokine which are elevated when there is ongoing inflammation in the CNS and correlate with increased disease activity in MS (Paper II). We have found risk alleles that are present together in MS cases more often than expected, thus being said to interact and likely to be present in the same sufficient cause for MS (Paper III). We have identified several risk loci that are eQTLs, some of which were replicated from previous studies, while several are new (Paper IV). In this thesis, I have only been able to speculate about functional roles in MS pathogenesis of candidate genes found through the studies, however I have not yet had the possibility to test their functions experimentally. This will be the subject of my future perspectives.

Why is it so important to understand disease mechanisms? Is it not enough that there are quite efficient treatments for MS today, even though we do not know exactly why they work well? Unfortunately the treatments do not work well for all persons with this highly heterogeneous disease, and are moreover associated with increased risk for serious adverse events. By studying disease mechanisms we hope to find new, personalized and more effectively targeted treatments, as well as ways to prevent disease.

We are still far from understanding how, and through which functions genetic variation contributes to MS, but hopefully the results presented in this thesis may be of help for future studies in this direction.

6 FUTURE PERSPECTIVES

There are many things that remain to be analyzed in the data that I have worked with in this thesis, and there are even more analyses and experiments that I would like to do as a follow-up to my findings.

As I mentioned earlier, further genetic studies of the *IL21* region do not seem to be of interest at this stage given the knowledge that we have about genetic influences in MS, although it cannot be excluded that we would go back and explore its function in a subgroup of patients in the future.

To follow up the study in Paper II, it would in the first place be of interest to analyze levels of CXCL13 and other potential CSF biomarkers in relation to an updated list of MS risk loci in an extended cohort with a greater number of individuals included. In parallel, our RNA-seq data could be used to evaluate whether differences in expression of the *CXCL13* gene can be observed in PBMCs in relation to genotypes from the ImmunoChip data. It has been shown that CXCL13 in the CSF is mainly produced by infiltrating monocytes to the CNS, and probably even more if they are differentiated into macrophages (110), therefore an eQTL study in monocytes (preferably purified from the CSF) and macrophages would be of great interest.

The findings from Paper III should be replicated in an independent and larger cohort, which is especially important in order to validate the suggestive interaction between *HLA-DRB1*15:01* and the rs6677309 A allele, since the reference group (negative for both risk alleles) was very small. The results in Paper III can be followed up by studying expression levels of HLA-DRB1 and other HLA class II molecules together with the expression of CD58 in relation to the SNP genotypes, and evaluate functional consequences such as T cell proliferation and induction of T_{regs}.

To follow up on the results presented in Paper IV, most importantly, the eQTLs should be validated in different cell types. Subtypes of immune cells are of primary interest, and most easy to access. However, other tissues, in particular cells in the CNS, should not be neglected. These tissues are obviously very difficult to access for human studies and therefore transgenic animal models will be useful. The animal models should be used to

evaluate normal functions of the candidate genes, as well as their functions in disease models for MS, such as the EAE.

The fact that we have not detected eQTL effects for some genes that have been reported by others, such as the CD58 gene or CYP27B1, underscores the importance of future cell type-specific eQTL analyses, and also that other genes, with less significant expression associations to the MS SNPs in our data, should be considered in future analyses. The RNA-seq data produced for Paper IV can also be used to study how MS variants influence the expression of transcript isoforms by quantification of exon expression.

7 ACKNOWLEDGEMENTS

Ingrid Kockum my main supervisor, you are a very special person. It has been a true privilege to be your student. You have a great knowledge that you generously share, and with which you help many persons in their work. Your kindness always shines through you. Thank you for everything you have taught me, for always making sure you find time to help when needed, for your patience, and for giving me the tools for working more and more independently. I wish you and your wonderful family all the best!

Tomas Olsson, my co-supervisor, you are the "dad" of the Neuroimmunology Unit. Thank you for giving me the opportunity to work in the Unit, for promoting an open and nice atmosphere, for your sense of humor, for being encouraging and enthusiastic about solving "the MS riddle".

Mohsen Khademi, my co-supervisor, thank you for everything you have taught me about gene expression analysis, biomarkers, MS patients and clinical data, for your endless patience in answering all my questions, for all the technical and experimental advice, for always being a gentleman. I also want to thank you for all the great work you do as a lab manager.

Rachel Fisher, my external mentor, thank you for your interest and support throughout the years before and during my PhD. Our chats have given me very good insights, which for sure will help to shape my future career plans. I hope we will keep in contact!

My co-authors:

Tojo James, thank you for the endless hours spent on coding and coding and recoding and coding... while working on that endless project. We have both learnt a lot during this RNA-seq journey. A special thanks also for help with running the interaction script on Athena – it's a hard life to be an informatician among people like us… Thank you for feeding us with chocolate and good luck with your thesis – you have a great future ahead of you! **Iza Lima Bomfim**, always so helpful and interested! Thank you for your smiling support, for helping me out with so many things, especially for explaining statistics. Your scientific curiousness and enthusiasm is truly inspiring! **Emilie Sundqvist,** thanks for the time we were working together, for being caring and supportive, and for always coming up with some good practical advice! **David Gomez Cabrero,** thanks for not killing me (=thanks for your patience) for my repeated questions on statistics… It has been great to work with you and I have learnt a lot - thanks! **Mikael Huss and Maya Brandi,** our colleagues from the SciLife Lab. Thank you for the collaboration! **Rita Nohra,** thank you for nice chats way back in time.

Good luck with your job! **Maja Jagodic,** thanks for being so nice and encouraging, **Fredrik Piehl, Jan Hillert, Lars Alfredsson** and **Jesper Tegnér,** thank you for commenting the manuscripts and for interesting input.

A special thanks to **Venus Azhary** for helping out with TaqMan genotypings (I wouldn't have made it without you!), for the work you do for all of us in the Unit, for promoting a nice working atmosphere!
Sahl Bedri, thanks for your work with the stimulation project. It didn't become a publication but we shall not give up! Let's see what will become of your favorite gene…
My great roomies:
Magnus Lekman, the interaction manuscript in this thesis would not have existed if it wasn't for your initiative and endless hours spent on the struggle to get through with the "jeira/jeisa/geisa/whatever…script". Thank you so much for this, for your friendliness and for all the help that you have always been willing to give in spite of having an enormous workload. I wish you all the best! **Sandra Gyllenberg,** thank you for your help and support, especially during thesis writing. Thank you for becoming a friend and for cheering up the working environment with funny stories about your wonderful children. You are a strong woman who knows what is important in life. **Nada Abdelmaggid,** I'm happy you moved to our room! Thanks for all good advice and caring support, and thanks for managing the labmeeting list! **Pernilla Stridh,** thanks for very interesting and inspiring discussions on science and life, for your wisdom, good advice and help. **Cecilia Dominguez**, we miss you in our room and lab! Thank you for nice chats, for creating a nice and warm atmosphere around you. **Lara,** our ex-roomie, it was great to have you in our writing room, mon amie française, and a pity that it didn't last long!
People in the Neuroimmunology Unit: **Petra, Marie, Brinda, Sabrina, Rasmus, Shahin, Hannes, Karl, André, Cynthia, Rux, Lou, Sreeni, Hanna, Clas**. **Bob** and his group: **Roham, Andreas, Sohel, Harald, Xing-Mei.** Former members of the Neuroimmunology Unit, especially **Mélanie, Ame, Alan, Rickard, Faiez, Patrick, Johan.** To all of you: Thanks for help, friendliness, chats and discussions on science and life.

Colleagues from the MS group, especially **Jenny, Malin, Kerstin, Christina, Elin** and **Andrius.**
Colleagues from the Unit for Computational Medicine, especially **Gilad** for sharing your knowledge and for interesting discussions, and **Francesco** for advice on heatmaps.
Colleagues from the Rheumatology Unit, particularly for interesting discussions on science and life with our human genetics fellows: **Leonid, Lina** (I'm especially grateful for your

friendship!), **Klementy, Espen.** The Polish girls, **Kinga** and **Agnes,** thanks for nice chats and good luck with your projects.

Core facility staff at KI gene, **Annika, Selim** and **Anna-Lee,** and at the Mutation Analysis Facility, **Kristina** and **Aida.**
Cleaning staff, especially **Asmelash** and **Nimo.** Autoclaving staff, previously **Catherina** and later **Edna.**

Ingeborg van der Ploeg for introducing me to the world of research by letting me perform my "PRAO" in your lab when I was 14, and for all good advice throughout the years.

My whole wonderful family, all my friends, present and former students from Ateneum: I will not write the whole list with names of all the persons that are important to me, and who have contributed to this thesis by helping me so much, supporting me and accompanying me on my way – it would fill a whole book!! You know who you are, and you are all in my heart.

I will only make one exception, my parents:
Mamusiu, nie mogę nawet wyrazić w słowach ile Ci zawdzięczam. Twoja milosc i Twoja walka w życiu są dla mnie ciagłą inspiracją podporą. **Pappa**, jag vet att du varit med mig hela tiden. Tack för att du varit med och gett mig livet, tack för din kärlek.

Finally, I'm thanking **God** for EVERYTHING.

8 REFERENCES

1. Kumar DR, Aslinia F, Yale SH, Mazza JJ. Jean-Martin Charcot: the father of neurology. Clinical medicine & research. 2011;9(1):46-9.

2. Noseworthy JH, Lucchinetti C, Rodriguez M, Weinshenker BG. Multiple sclerosis. The New England journal of medicine. 2000;343(13):938-52.

3. Ahlgren C, Oden A, Lycke J. High nationwide prevalence of multiple sclerosis in Sweden. Multiple sclerosis. 2011;17(8):901-8.

4. Gronlie SA, Myrvoll E, Hansen G, Gronning M, Mellgren SI. Multiple sclerosis in North Norway, and first appearance in an indigenous population. Journal of neurology. 2000;247(2):129-33.

5. Pugliatti M, Sotgiu S, Solinas G, Castiglia P, Rosati G. Multiple sclerosis prevalence among Sardinians: further evidence against the latitude gradient theory. Neurological sciences : official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology. 2001;22(2):163-5.

6. Westerlind H, Ramanujam R, Uvehag D, Kuja-Halkola R, Boman M, Bottai M, et al. Modest familial risks for multiple sclerosis: a registry-based study of the population of Sweden. Brain : a journal of neurology. 2014;137(Pt 3):770-8.

7. Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. Annals of neurology. 2011;69(2):292-302.

8. Compston A, Coles A. Multiple sclerosis. Lancet. 2008;372(9648):1502-17.

9. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology. 1983;33(11):1444-52.

10. Dobson R, Ramagopalan S, Davis A, Giovannoni G. Cerebrospinal fluid oligoclonal bands in multiple sclerosis and clinically isolated syndromes: a meta-analysis of prevalence, prognosis and effect of latitude. Journal of neurology, neurosurgery, and psychiatry. 2013;84(8):909-14.

11. Prineas JW, Graham JS. Multiple sclerosis: capping of surface immunoglobulin G on macrophages engaged in myelin breakdown. Annals of neurology. 1981;10(2):149-58.

12. Babbe H, Roers A, Waisman A, Lassmann H, Goebels N, Hohlfeld R, et al. Clonal expansions of CD8(+) T cells dominate the T cell infiltrate in active multiple sclerosis lesions as shown by micromanipulation and single cell polymerase chain reaction. The Journal of experimental medicine. 2000;192(3):393-404.

13. Wucherpfennig KW, Newcombe J, Li H, Keddy C, Cuzner ML, Hafler DA. Gamma delta T-cell receptor repertoire in acute multiple sclerosis lesions. Proceedings of the National Academy of Sciences of the United States of America. 1992;89(10):4588-92.

14. Nylander A, Hafler DA. Multiple sclerosis. The Journal of clinical investigation. 2012;122(4):1180-8.

15. Fletcher JM, Lalor SJ, Sweeney CM, Tubridy N, Mills KH. T cells in multiple sclerosis and experimental autoimmune encephalomyelitis. Clinical and experimental immunology. 2010;162(1):1-11.

16. Brucklacher-Waldert V, Stuerner K, Kolster M, Wolthausen J, Tolosa E. Phenotypical and functional characterization of T helper 17 cells in multiple sclerosis. Brain : a journal of neurology. 2009;132(Pt 12):3329-41.

17. Steinman L. A brief history of T(H)17, the first major revision in the T(H)1/T(H)2 hypothesis of T cell-mediated tissue damage. Nature medicine. 2007;13(2):139-45.

18. Viglietta V, Baecher-Allan C, Weiner HL, Hafler DA. Loss of functional suppression by CD4+CD25+ regulatory T cells in patients with multiple sclerosis. The Journal of experimental medicine. 2004;199(7):971-9.

19. Haas J, Hug A, Viehover A, Fritzsching B, Falk CS, Filser A, et al. Reduced suppressive effect of CD4+CD25high regulatory T cells on the T cell immune response against myelin oligodendrocyte glycoprotein in patients with multiple sclerosis. European journal of immunology. 2005;35(11):3343-52.

20. Bar-Or A, Fawaz L, Fan B, Darlington PJ, Rieger A, Ghorayeb C, et al. Abnormal B-cell cytokine responses a trigger of T-cell-mediated disease in MS? Annals of neurology. 2010;67(4):452-61.

21. Bankoti J, Apeltsin L, Hauser SL, Allen S, Albertolle ME, Witkowska HE, et al. In multiple sclerosis, oligoclonal bands connect to peripheral B-cell responses. Annals of neurology. 2014;75(2):266-76.

22. Hauser SL, Waubant E, Arnold DL, Vollmer T, Antel J, Fox RJ, et al. B-cell depletion with rituximab in relapsing-remitting multiple sclerosis. The New England journal of medicine. 2008;358(7):676-88.

23. Kappos L, Li D, Calabresi PA, O'Connor P, Bar-Or A, Barkhof F, et al. Ocrelizumab in relapsing-remitting multiple sclerosis: a phase 2, randomised, placebo-controlled, multicentre trial. Lancet. 2011;378(9805):1779-87.

24. Serafini B, Rosicarelli B, Magliozzi R, Stigliano E, Aloisi F. Detection of ectopic B-cell follicles with germinal centers in the meninges of patients with secondary progressive multiple sclerosis. Brain Pathol. 2004;14(2):164-74.

25. Franciotta D, Salvetti M, Lolli F, Serafini B, Aloisi F. B cells and multiple sclerosis. Lancet neurology. 2008;7(9):852-8.

26. Chastain EM, Miller SD. Molecular mimicry as an inducing trigger for CNS autoimmune demyelinating disease. Immunological reviews. 2012;245(1):227-38.

27. Bo L, Geurts JJ, van der Valk P, Polman C, Barkhof F. Lack of correlation between cortical demyelination and white matter pathologic changes in multiple sclerosis. Archives of neurology. 2007;64(1):76-80.

28. Lassmann H, van Horssen J, Mahad D. Progressive multiple sclerosis: pathology and pathogenesis. Nature reviews Neurology. 2012;8(11):647-56.

29. Group PS, the University of British Columbia MSMRIAG. PRISMS-4: Long-term efficacy of interferon-beta-1a in relapsing MS. Neurology. 2001;56(12):1628-36.

30. Kappos L, Freedman MS, Polman CH, Edan G, Hartung HP, Miller DH, et al. Long-term effect of early treatment with interferon beta-1b after a first clinical event suggestive of multiple sclerosis: 5-year active treatment extension of the phase 3 BENEFIT trial. Lancet neurology. 2009;8(11):987-97.

31. Dhib-Jalbut S, Marks S. Interferon-beta mechanisms of action in multiple sclerosis. Neurology. 2010;74 Suppl 1:S17-24.

32. Carter NJ, Keating GM. Glatiramer acetate: a review of its use in relapsing-remitting multiple sclerosis and in delaying the onset of clinically definite multiple sclerosis. Drugs. 2010;70(12):1545-77.

33. Polman CH, O'Connor PW, Havrdova E, Hutchinson M, Kappos L, Miller DH, et al. A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis. The New England journal of medicine. 2006;354(9):899-910.

34. Kappos L, Bates D, Edan G, Eraksoy M, Garcia-Merino A, Grigoriadis N, et al. Natalizumab treatment for multiple sclerosis: updated recommendations for patient selection and monitoring. Lancet neurology. 2011;10(8):745-58.

35. Rothman KJG, S.;Lash, T.L. Modern Epidemiology: Lippincott Williams & Wilkins; 2008.

36. Ascherio A, Munger KL, Simon KC. Vitamin D and multiple sclerosis. Lancet neurology. 2010;9(6):599-612.

37. Baarnhielm M, Hedstrom AK, Kockum I, Sundqvist E, Gustafsson SA, Hillert J, et al. Sunlight is associated with decreased multiple sclerosis risk: no interaction with human leukocyte antigen-DRB1*15. European journal of neurology : the official journal of the European Federation of Neurological Societies. 2012;19(7):955-62.

38. Hernan MA, Olek MJ, Ascherio A. Cigarette smoking and incidence of multiple sclerosis. American journal of epidemiology. 2001;154(1):69-74.

39. Hedstrom AK, Baarnhielm M, Olsson T, Alfredsson L. Tobacco smoking, but not Swedish snuff use, increases the risk of multiple sclerosis. Neurology. 2009;73(9):696-701.

40. Sundqvist E, Sundstrom P, Linden M, Hedstrom AK, Aloisi F, Hillert J, et al. Epstein-Barr virus and multiple sclerosis: interaction with HLA. Genes and immunity. 2012;13(1):14-20.

41. Thacker EL, Mirzaei F, Ascherio A. Infectious mononucleosis and risk for multiple sclerosis: a meta-analysis. Annals of neurology. 2006;59(3):499-503.

42. Munger KL, Chitnis T, Ascherio A. Body size and risk of MS in two cohorts of US women. Neurology. 2009;73(19):1543-50.

43. Hedstrom AK, Olsson T, Alfredsson L. High body mass index before age 20 is associated with increased risk for multiple sclerosis in both men and women. Multiple sclerosis. 2012;18(9):1334-6.

44. Hedstrom AK, Lima Bomfim I, Barcellos L, Gianfrancesco M, Schaefer C, Kockum I, et al. Interaction between adolescent obesity and HLA risk genes in the etiology of multiple sclerosis. Neurology. 2014;82(10):865-72.

45. Harbo HF, Utsi E, Lorentzen AR, Kampman MT, Celius EG, Myhr KM, et al. Low frequency of the disease-associated DRB1*15-DQB1*06 haplotype may contribute to the low prevalence of multiple sclerosis in Sami. Tissue antigens. 2007;69(4):299-304.

46. Harbo HF, Riccio ME, Lorentzen AR, Utsi E, Myhr KM, Mellgren SI, et al. Norwegian Sami differs significantly from other Norwegians according to their HLA profile. Tissue antigens. 2010;75(3):207-17.

47. Marrosu MG, Muntoni F, Murru MR, Spinicci G, Pischedda MP, Goddi F, et al. Sardinian multiple sclerosis is associated with HLA-DR4: a serologic and molecular analysis. Neurology. 1988;38(11):1749-53.

48. Muntoni F, Murru MR, Costa G, Congia M, Cucca F, Cossu P, et al. Different HLA DR2-DQw1 haplotypes in Sardinian and northern Italian populations: implications for multiple sclerosis susceptibility. Tissue antigens. 1991;38(1):34-6.

49. Marrosu MG, Murru R, Murru MR, Costa G, Zavattari P, Whalen M, et al. Dissection of the HLA association with multiple sclerosis in the founder isolated population of Sardinia. Human molecular genetics. 2001;10(25):2907-16.

50. Arlt MF, Wilson TE, Glover TW. Replication stress and mechanisms of CNV formation. Current opinion in genetics & development. 2012;22(3):204-10.

51. Jersild C, Fog T. Histocompatibility (HL-A) antigens associated with multiple sclerosis. Acta neurologica Scandinavica Supplementum. 1972;51:377.

52. Jersild C, Svejgaard A, Fog T. HL-A antigens and multiple sclerosis. Lancet. 1972;1(7762):1240-1.

53. Cazzullo CL, Smeraldi E. HL-A antigens and multiple sclerosis. Lancet. 1972;2(7774):429-30.

54. Horton R, Wilming L, Rand V, Lovering RC, Bruford EA, Khodiyar VK, et al. Gene map of the extended human MHC. Nature reviews Genetics. 2004;5(12):889-99.

55. Hillert J, Olerup O. Multiple sclerosis is associated with genes within or close to the HLA-DR-DQ subregion on a normal DR15,DQ6,Dw2 haplotype. Neurology. 1993;43(1):163-8.

56. International Multiple Sclerosis Genetics C, Wellcome Trust Case Control C, Sawcer S, Hellenthal G, Pirinen M, Spencer CC, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature. 2011;476(7359):214-9.

57. Patsopoulos NA, Barcellos LF, Hintzen RQ, Schaefer C, van Duijn CM, Noble JA, et al. Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects. PLoS genetics. 2013;9(11):e1003926.

58. Oksenberg JR, Barcellos LF, Cree BA, Baranzini SE, Bugawan TL, Khan O, et al. Mapping multiple sclerosis susceptibility to the HLA-DR locus in African Americans. American journal of human genetics. 2004;74(1):160-7.

59. Fogdell-Hahn A, Ligers A, Gronning M, Hillert J, Olerup O. Multiple sclerosis: a modifying influence of HLA class I genes in an HLA class II associated autoimmune disease. Tissue antigens. 2000;55(2):140-8.

60. Brynedal B, Duvefelt K, Jonasdottir G, Roos IM, Akesson E, Palmgren J, et al. HLA-A confers an HLA-DRB1 independent influence on the risk of multiple sclerosis. PloS one. 2007;2(7):e664.

61. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, et al. Complement factor H polymorphism in age-related macular degeneration. Science. 2005;308(5720):385-9.

62. Barsh GS, Copenhaver GP, Gibson G, Williams SM. Guidelines for genome-wide association studies. PLoS genetics. 2012;8(7):e1002812.

63. International Multiple Sclerosis Genetics C, Hafler DA, Compston A, Sawcer S, Lander ES, Daly MJ, et al. Risk alleles for multiple sclerosis identified by a genomewide study. The New England journal of medicine. 2007;357(9):851-62.

64. Patsopoulos NA, Bayer Pharma MSGWG, Steering Committees of Studies Evaluating I-b, a CCRA, Consortium AN, GeneMsa, et al. Genome-wide meta-analysis identifies novel multiple sclerosis susceptibility loci. Annals of neurology. 2011;70(6):897-912.

65. Parkes M, Cortes A, van Heel DA, Brown MA. Genetic insights into common pathways and complex relationships among immune-mediated diseases. Nature reviews Genetics. 2013;14(9):661-73.

66. International Multiple Sclerosis Genetics C, Beecham AH, Patsopoulos NA, Xifara DK, Davis MF, Kemppinen A, et al. Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. Nature genetics. 2013;45(11):1353-60.

67. Zhou Y, Simpson S, Jr., Holloway AF, Charlesworth J, van der Mei I, Taylor BV. The potential role of epigenetic modifications in the heritability of multiple sclerosis. Multiple sclerosis. 2014;20(2):135-40.

68. Lawson HA, Cheverud JM, Wolf JB. Genomic imprinting and parent-of-origin effects on complex traits. Nature reviews Genetics. 2013;14(9):609-17.

69. Hedstrom AK, Sundqvist E, Baarnhielm M, Nordin N, Hillert J, Kockum I, et al. Smoking and two human leukocyte antigen genes interact to increase the risk for multiple sclerosis. Brain : a journal of neurology. 2011;134(Pt 3):653-64.

70. Bateson W. Mendel's Principles of Heredity: Camebridge [Eng.] University Press; 1909.

71. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science. 1995;270(5235):467-70.

72. Kemppinen AK, Kaprio J, Palotie A, Saarela J. Systematic review of genome-wide expression studies in multiple sclerosis. BMJ open. 2011;1(1):e000053.

73. Fehrmann RS, Jansen RC, Veldink JH, Westra HJ, Arends D, Bonder MJ, et al. Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. PLoS genetics. 2011;7(8):e1002197.

74. Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H, et al. Common regulatory variation impacts gene expression in a cell type-dependent manner. Science. 2009;325(5945):1246-50.

75. Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, Beazley C, et al. Population genomics of human gene expression. Nature genetics. 2007;39(10):1217-24.

76. Pickrell JK, Marioni JC, Pai AA, Degner JF, Engelhardt BE, Nkadori E, et al. Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature. 2010;464(7289):768-72.

77. Montgomery SB, Sammeth M, Gutierrez-Arcelus M, Lach RP, Ingle C, Nisbett J, et al. Transcriptome genetics using second generation sequencing in a Caucasian population. Nature. 2010;464(7289):773-7.

78. Lappalainen T, Sammeth M, Friedlander MR, t Hoen PA, Monlong J, Rivas MA, et al. Transcriptome and genome sequencing uncovers functional variation in humans. Nature. 2013;501(7468):506-11.

79. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews Genetics. 2009;10(1):57-63.

80. Biomarkers Definitions Working G. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clinical pharmacology and therapeutics. 2001;69(3):89-95.

81. Braitch M, Nunan R, Niepel G, Edwards LJ, Constantinescu CS. Increased osteopontin levels in the cerebrospinal fluid of patients with multiple sclerosis. Archives of neurology. 2008;65(5):633-5.

82. Sato W, Tomita A, Ichikawa D, Lin Y, Kishida H, Miyake S, et al. CCR2(+)CCR5(+) T cells produce matrix metalloproteinase-9 and osteopontin in the pathogenesis of multiple sclerosis. J Immunol. 2012;189(10):5057-65.

83. Teunissen CE, Khalil M. Neurofilaments as biomarkers in multiple sclerosis. Multiple sclerosis. 2012;18(5):552-6.

84. Khademi M, Dring AM, Gilthorpe JD, Wuolikainen A, Al Nimer F, Harris RA, et al. Intense inflammation and nerve damage in early multiple sclerosis subsides at older age: a reflection by cerebrospinal fluid biomarkers. PloS one. 2013;8(5):e63172.

85. Brettschneider J, Czerwoniak A, Senel M, Fang L, Kassubek J, Pinkhardt E, et al. The chemokine CXCL13 is a prognostic marker in clinically isolated syndrome (CIS). PloS one. 2010;5(8):e11986.

86. Khademi M, Kockum I, Andersson ML, Iacobaeus E, Brundin L, Sellebjerg F, et al. Cerebrospinal fluid CXCL13 in multiple sclerosis: a suggestive prognostic marker for the disease course. Multiple sclerosis. 2011;17(3):335-43.

87. Consortium EP, Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74.

88. Piehl F, Holmen C, Hillert J, Olsson T. Swedish natalizumab (Tysabri) multiple sclerosis surveillance study. Neurological sciences : official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology. 2011;31 Suppl 3:289-93.

89. Poser CM, Paty DW, Scheinberg L, McDonald WI, Davis FA, Ebers GC, et al. New diagnostic criteria for multiple sclerosis: guidelines for research protocols. Annals of neurology. 1983;13(3):227-31.

90. Polman CH, Reingold SC, Edan G, Filippi M, Hartung HP, Kappos L, et al. Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria". Annals of neurology. 2005;58(6):840-6.

91. Hedstrom AK, Hillert J, Olsson T, Alfredsson L. Alcohol as a modifiable lifestyle factor affecting multiple sclerosis risk. JAMA neurology. 2014;71(3):300-5.

92. Olerup O, Zetterquist H. HLA-DR typing by PCR amplification with sequence-specific primers (PCR-SSP) in 2 hours: an alternative to serological DR typing in clinical practice including donor-recipient matching in cadaveric transplantation. Tissue antigens. 1992;39(5):225-35.

93. Dilthey A, Leslie S, Moutsianas L, Shen J, Cox C, Nelson MR, et al. Multi-population classical HLA type imputation. PLoS computational biology. 2013;9(2):e1002877.

94. Roos IM, Kockum I, Hillert J. The interleukin 23 receptor gene in multiple sclerosis: a case-control study. Journal of neuroimmunology. 2008;194(1-2):173-80.

95. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic acids research. 2001;29(9):e45.

96. Pfaffl MW, Horgan GW, Dempfle L. Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic acids research. 2002;30(9):e36.

97. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001;25(4):402-8.

98. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome biology. 2002;3(7):RESEARCH0034.

99. Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome biology. 2007;8(2):R19.

100. Anders S PP, Huber W. HTSeq — A Python framework to work with high-throughput sequencing data. 2014.

101. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome biology. 2010;11(3):R25.

102. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics. 2007;81(3):559-75.

103. Team RDC. R: A language and environment for statistical computing. R Foundation for Statistical Computing Vienna, Austria2011. Available from: http://www.R-project.org/.

104. Ding B, Kallberg H, Klareskog L, Padyukov L, Alfredsson L. GEIRA: gene-environment and gene-gene interaction research application. European journal of epidemiology. 2011;26(7):557-61.

105. Li J, Tibshirani R. Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data. Statistical methods in medical research. 2013;22(5):519-36.

106. Nohra R, Beyeen AD, Guo JP, Khademi M, Sundqvist E, Hedreul MT, et al. RGMA and IL21R show association with experimental inflammation and multiple sclerosis. Genes and immunity. 2010;11(4):279-93.

107. Sawalha AH, Kaufman KM, Kelly JA, Adler AJ, Aberle T, Kilpatrick J, et al. Genetic association of interleukin-21 polymorphisms with systemic lupus erythematosus. Annals of the rheumatic diseases. 2008;67(4):458-61.

108. Webb R, Merrill JT, Kelly JA, Sestak A, Kaufman KM, Langefeld CD, et al. A polymorphism within IL21R confers risk for systemic lupus erythematosus. Arthritis and rheumatism. 2009;60(8):2402-7.

109. Fedetz M, Ndagire D, Fernandez O, Leyva L, Guerrero M, Arnal C, et al. Multiple sclerosis association study with the TENR-IL2-IL21 region in a Spanish population. Tissue antigens. 2009;74(3):244-7.

110. Krumbholz M, Theil D, Cepok S, Hemmer B, Kivisakk P, Ransohoff RM, et al. Chemokines in multiple sclerosis: CXCL12 and CXCL13 up-regulation is differentially linked to CNS immune cell recruitment. Brain : a journal of neurology. 2006;129(Pt 1):200-11.

111. Alvarez E, Piccio L, Mikesell RJ, Klawiter EC, Parks BJ, Naismith RT, et al. CXCL13 is a biomarker of inflammation in multiple sclerosis, neuromyelitis optica, and other neurological conditions. Multiple sclerosis. 2013;19(9):1204-8.

112. Rupprecht TA, Pfister HW, Angele B, Kastenbauer S, Wilske B, Koedel U. The chemokine CXCL13 (BLC): a putative diagnostic marker for neuroborreliosis. Neurology. 2005;65(3):448-50.

113. Courtioux B, Pervieux L, Vatunga G, Marin B, Josenando T, Jauberteau-Marchan MO, et al. Increased CXCL-13 levels in human African trypanosomiasis meningo-encephalitis. Tropical medicine & international health : TM & IH. 2009;14(5):529-34.

114. Barcellos LF, Oksenberg JR, Begovich AB, Martin ER, Schmidt S, Vittinghoff E, et al. HLA-DR2 dose effect on susceptibility to multiple sclerosis and influence on disease course. American journal of human genetics. 2003;72(3):710-6.

115. Sellebjerg F, Jensen J, Madsen HO, Svejgaard A. HLA DRB1*1501 and intrathecal inflammation in multiple sclerosis. Tissue antigens. 2000;55(4):312-8.

116. Moratz C, Kang VH, Druey KM, Shi CS, Scheschonka A, Murphy PM, et al. Regulator of G protein signaling 1 (RGS1) markedly impairs Gi alpha signaling responses of B lymphocytes. J Immunol. 2000;164(4):1829-38.

117. Moratz C, Hayman JR, Gu H, Kehrl JH. Abnormal B-cell responses to chemokines, disturbed plasma cell localization, and distorted immune tissue architecture in Rgs1-/- mice. Molecular and cellular biology. 2004;24(13):5767-75.

118. Filippi A, Tiso N, Deflorian G, Zecchin E, Bortolussi M, Argenton F. The basic helix-loop-helix olig3 establishes the neural plate boundary of the trunk and is necessary for development of the dorsal spinal cord. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(12):4377-82.

119. Wertz IE, O'Rourke KM, Zhou H, Eby M, Aravind L, Seshagiri S, et al. De-ubiquitination and ubiquitin ligase domains of A20 downregulate NF-kappaB signalling. Nature. 2004;430(7000):694-9.

120. Kool M, van Loo G, Waelput W, De Prijck S, Muskens F, Sze M, et al. The ubiquitin-editing protein A20 prevents dendritic cell activation, recognition of apoptotic cells, and systemic autoimmunity. Immunity. 2011;35(1):82-96.

121. Kristjansdottir G, Sandling JK, Bonetti A, Roos IM, Milani L, Wang C, et al. Interferon regulatory factor 5 (IRF5) gene variants are associated with multiple sclerosis in three distinct populations. Journal of medical genetics. 2008;45(6):362-9.

122. Friese MA, Jakobsen KB, Friis L, Etzensperger R, Craner MJ, McMahon RM, et al. Opposing effects of HLA class I molecules in tuning autoreactive CD8+ T cells in multiple sclerosis. Nature medicine. 2008;14(11):1227-35.

123. Burfoot RK, Jensen CJ, Field J, Stankovich J, Varney MD, Johnson LJ, et al. SNP mapping and candidate gene sequencing in the class I region of the HLA complex: searching for multiple sclerosis susceptibility genes in Tasmanians. Tissue antigens. 2008;71(1):42-50.

124. De Jager PL, Baecher-Allan C, Maier LM, Arthur AT, Ottoboni L, Barcellos L, et al. The role of the CD58 locus in multiple sclerosis. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(13):5264-9.

125. Handel AE, Handunnetthi L, Berlanga AJ, Watson CT, Morahan JM, Ramagopalan SV. The effect of single nucleotide polymorphisms from genome wide association studies in multiple sclerosis on gene expression. PloS one. 2010;5(4):e10142.

126. Kanner SB, Damle NK, Blake J, Aruffo A, Ledbetter JA. CD2/LFA-3 ligation induces phospholipase-C gamma 1 tyrosine phosphorylation and regulates CD3 signaling. J Immunol. 1992;148(7):2023-9.

127. Sanchez-Madrid F, Krensky AM, Ware CF, Robbins E, Strominger JL, Burakoff SJ, et al. Three distinct antigens associated with human T-lymphocyte-mediated cytolysis: LFA-1, LFA-2, and LFA-3. Proceedings of the National Academy of Sciences of the United States of America. 1982;79(23):7489-93.

128. Alcina A, Abad-Grau Mdel M, Fedetz M, Izquierdo G, Lucas M, Fernandez O, et al. Multiple sclerosis risk variant HLA-DRB1*1501 associates with high expression of DRB1 gene in different human populations. PloS one. 2012;7(1):e29819.

129. Folkersen L, van't Hooft F, Chernogubova E, Agardh HE, Hansson GK, Hedin U, et al. Association of genetic risk variants with expression of proximal genes identifies novel susceptibility genes for cardiovascular disease. Circulation Cardiovascular genetics. 2010;3(4):365-73.

130. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science. 2012;337(6099):1190-5.

131. Ryan EJ, Marshall AJ, Magaletti D, Floyd H, Draves KE, Olson NE, et al. Dendritic cell-associated lectin-1: a novel dendritic cell-associated, C-type lectin-like molecule enhances T cell secretion of IL-4. J Immunol. 2002;169(10):5638-48.

132. Bouwmeester T, Bauch A, Ruffner H, Angrand PO, Bergamini G, Croughton K, et al. A physical and functional map of the human TNF-alpha/NF-kappa B signal transduction pathway. Nature cell biology. 2004;6(2):97-105.

133. Li S, Wang L, Berman M, Kong YY, Dorf ME. Mapping a dynamic innate immunity protein interaction network regulating type I interferon production. Immunity. 2011;35(3):426-40.

134. Carreras-Sureda A, Cantero-Recasens G, Rubio-Moscardo F, Kiefer K, Peinelt C, Niemeyer BA, et al. ORMDL3 modulates store-operated calcium entry and lymphocyte activation. Human molecular genetics. 2013;22(3):519-30.

135. Dixon-Salazar T, Silhavy JL, Marsh SE, Louie CM, Scott LC, Gururaj A, et al. Mutations in the AHI1 gene, encoding jouberin, cause Joubert syndrome with cortical polymicrogyria. American journal of human genetics. 2004;75(6):979-87.

136. Zhou LL, Zhao Y, Ringrose A, DeGeer D, Kennah E, Lin AE, et al. AHI-1 interacts with BCR-ABL and modulates BCR-ABL transforming activity and imatinib response of CML stem/progenitor cells. The Journal of experimental medicine. 2008;205(11):2657-71.

137. Carpentier M, Guillemette C, Bailey JL, Boileau G, Jeannotte L, DesGroseillers L, et al. Reduced fertility in male mice deficient in the zinc metallopeptidase NL1. Molecular and cellular biology. 2004;24(10):4428-37.

138. Ouimet T, Facchinetti P, Rose C, Bonhomme MC, Gros C, Schwartz JC. Neprilysin II: A putative novel metalloprotease and its isoforms in CNS and testis. Biochemical and biophysical research communications. 2000;271(3):565-70.

139. Huang JY, Hafez DM, James BD, Bennett DA, Marr RA. Altered NEP2 expression and activity in mild cognitive impairment and Alzheimer's disease. Journal of Alzheimer's disease : JAD. 2012;28(2):433-41.

140. De Jager PL, Jia X, Wang J, de Bakker PI, Ottoboni L, Aggarwal NT, et al. Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. Nature genetics. 2009;41(7):776-82.

141. Hoppenbrouwers IA, Aulchenko YS, Janssens AC, Ramagopalan SV, Broer L, Kayser M, et al. Replication of CD58 and CLEC16A as genome-wide significant risk genes for multiple sclerosis. Journal of human genetics. 2009;54(11):676-80.

142. Jakkula E, Leppa V, Sulonen AM, Varilo T, Kallio S, Kemppinen A, et al. Genome-wide association study in a high-risk isolate for multiple sclerosis reveals associated variants in STAT3 gene. American journal of human genetics. 2010;86(2):285-91.